Theses and Dissertations   1. Thesis and Dissertation Collection, all items

2005-06

The Surface Warfare community's 360-degree feedback pilot program : a preliminary analysis and evaluation plan

Williams, James M.

Monterey, California. Naval Postgraduate School

http://hdl.handle.net/10945/1919

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

THE SURFACE WARFARE COMMUNITY'S 360-DEGREE FEEDBACK PILOT PROGRAM: A PRELIMINARY ANALYSIS AND EVALUATION PLAN

by

James M. Williams

June 2005

Advisor:	Susan Hocevar
Associate Advisor:	Alice Crawford

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE		*Form Approved OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE June 2005	3. REPORT TYPE AND DATES COVERED Master's Thesis	
4. TITLE AND SUBTITLE: The Surface Warfare Community's 360-Degree Feedback Pilot Program: A Preliminary Analysis and Evaluation Plan		**5. FUNDING NUMBERS**	
6. AUTHOR(S) James M. Williams			

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
   Naval Postgraduate School
   Monterey, CA 93943-5000
		8. PERFORMING ORGANIZATION REPORT NUMBER	

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A		**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**	

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited	12b. DISTRIBUTION CODE A

13. ABSTRACT (maximum 200 words)

The system known as 360-degree feedback, also called multi-source or multi-rater feedback, is a development program that provides a recipient with feedback from supervisors, peers, and subordinates. There is currently no institutionalized, Navy-wide 360-degree feedback program for leadership development. Due to widespread civilian acceptance and to the success of the 360-degree program for the Navy's flag officers, the 2004 Surface Warfare Commanders Conference recommended a pilot program for 360-degree feedback be tested on a portion of the Surface Warfare Officer community. Results of the pilot program will be used to inform decisions on implementation of a Navy-wide 360-degree feedback program. The objectives of this thesis were to review the research evidence in the literature on the effectiveness and best practices of 360-degree programs and to identify general program evaluation techniques. The thesis then presents a conceptual analysis of the Navy pilot program and makes recommendations for modifications to the program based on comparisons with empirical research evidence and identified best practices of 360-degree programs. The thesis concludes by developing some guidelines and recommendations for a program evaluation plan that can be used to assess or revise the pilot program during and after its implementation.

14. SUBJECT TERMS 360-Degree Feedback, Multi-Source Feedback, Multi-Rater Feedback, Program Evaluation	15. NUMBER OF PAGES 87
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

THE SURFACE WARFARE COMMUNITY'S 360-DEGREE FEEDBACK PILOT PROGRAM: A PRELIMINARY ANALYSIS AND EVALUATION PLAN

James M. Williams
Lieutenant, United States Naval Reserve
B.B.A., University of Memphis, 1999

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF BUSINESS ADMINISTRATION

from the

**NAVAL POSTGRADUATE SCHOOL
June 2005**

Author: James M. Williams

Approved by: Susan Hocevar
                 Advisor

                 Alice Crawford
                 Associate Advisor

                 Douglas Brook
                 Dean, Graduate School of Business and Public Policy

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The system known as 360-degree feedback, also called multi-source or multi-rater feedback, is a development program that provides a recipient with feedback from supervisors, peers, and subordinates. There is currently no institutionalized, Navy-wide 360-degree feedback program for leadership development. Due to widespread civilian acceptance and to the success of the 360-degree program for the Navy's flag officers, the 2004 Surface Warfare Commanders Conference recommended a pilot program for 360-degree feedback be tested on a portion of the Surface Warfare Officer community. Results of the pilot program will be used to inform decisions on implementation of a Navy-wide 360-degree feedback program. The objectives of this thesis were to review the research evidence in the literature on the effectiveness and best practices of 360-degree programs and to identify general program evaluation techniques. The thesis then presents a conceptual analysis of the Navy pilot program and makes recommendations for modifications to the program based on comparisons with empirical research evidence and identified best practices of 360-degree programs. The thesis concludes by developing some guidelines and recommendations for a program evaluation plan that can be used to assess or revise the pilot program during and after its implementation.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. PURPOSE

The purpose of this research is to examine the effectiveness and best practices of 360-degree feedback programs in both the civilian and military communities. The intent is to compare the current Navy pilot program with available research and best practices, identify discrepancies, make recommendations for improvement, and provide a guideline for pilot program evaluation.

B. BACKGROUND

The system known as 360-degree feedback, also called multi-source or multi-rater feedback, is a development program that provides a recipient with feedback from supervisors, peers, and subordinates. The underlying theory of a 360-degree program is that there is variation in the ratings of different groups, and that this dissimilarity presents the recipient with meaningful information from different perspectives within the organization (LeBreton, Burgess, Kaiser, Atchley, and James, 2003).

The use of 360-degree programs in corporate America substantially increased during the 1990s (Brutus and Derayeh, 2002). Today 360-degree programs have achieved near-universal acceptance as leadership development tools, especially in Fortune 500 companies (Ghorpade, 2000).

There is currently no institutionalized, Navy-wide 360-degree feedback program for leadership development. Although the Navy strongly encourages mentoring for personal development, the only formal feedback process used Navy-wide is the current Fitness Report and Evaluation system, which is designed primarily for performance appraisal and provides only "top down" feedback on performance.

The 2004 Surface Warfare Commanders Conference recommended a pilot program for 360-degree feedback be tested on a portion of the Surface Warfare Officer community. The pilot is to be a sustained, three-year trial of 360-degree feedback administered to approximately five percent of Surface Warfare Officers. The main purpose of the pilot program is to determine effectiveness and feasibility of further Navy-wide implementation.

1

C.   RESEARCH OBJECTIVES

The primary research objectives are:

- To determine if 360-degree feedback programs are effective development tools.

- To identify best practices and lessons learned from civilian and military 360-degree feedback programs.

- To compare the Navy's 360-degree feedback pilot program to identified best practices and lessons learned.

- To provide a program evaluation guideline to assist the Navy in properly evaluating the effectiveness of the pilot program.

D.   SCOPE AND METHODOLOGY

The scope of this thesis is largely conceptual. The pilot program began in late 2004 and will continue through late 2007; therefore pilot data are not yet available for analysis. The thesis will present a conceptual analysis of the Navy pilot program as compared to empirical research and identified best practices and will also develop a framework for further program evaluation when pilot program empirical data are available.

The primary methodology for this research includes a literature review of empirical studies of both civilian and military 360-degree programs. Best practices, lessons learned, and program evaluation techniques are also identified through the literature review and personal interviews. Conclusions and recommendations for the Navy's pilot program are determined by comparing the current program plan with the identified best practices and lessons learned from the literature as well as established program evaluation techniques.

E.   EXPECTED BENEFITS

This thesis will provide the Navy with current knowledge regarding 360-degree program effectiveness, best practices, and overall program evaluation. This knowledge is

crucial for the Navy to properly analyze the design of the pilot program and accurately assess the costs and benefits of Navy-wide implementation of a 360-degree feedback program.

F.   THESIS ORGANIZATION

This thesis is partitioned into six chapters: Chapter II presents a brief history of 360-degree feedback use and a review of empirical data on the effectiveness of 360-degree programs as development tools. Chapter III presents a review of civilian and military program best practices and lessons learned in operating and enhancing the effectives of a 360-degree program. Chapter IV presents a thorough review of the Navy's 360-degree pilot program. Chapter V discusses program evaluation techniques in general and provides an analysis of the planned pilot program evaluation methods. Chapter VI presents conclusions and offers recommendations for adjustments to the pilot program.

THIS PAGE INTENTIONALLY LEFT BLANK

II.   360-DEGREE FEEDBACK

A.   INTRODUCTION

360-degree feedback, also called multi-source or multi-rater feedback, is a leadership performance evaluation and development program that uses assessments from superiors, peers, subordinates, and self to provide an individual a more thorough review of personal performance than is typically given in a traditional top-down assessment from a supervisor. The use of 360-degree programs in corporate America substantially increased in the 1990s to the point of near-universal acceptance in Fortune 500 companies (Ghorpade, 2000). This chapter presents a description and brief history of 360-degree program use and a detailed literature review of empirical studies that present contradictory findings on the effectiveness of 360-degree programs as development tools.

B.   DESCRIPTION OF 360-DEGREE FEEDBACK

Lepsinger and Lucia (1997) describe 360-degree feedback as a process where supervisors, peers, subordinates, and even customers provide perceptions about a person's behavior and the impact of that behavior as viewed from their various organizational perspectives. Downward feedback is provided by supervisors, upward feedback is provided by subordinates, and peer feedback is provided by individuals from the same organizational level as the feedback recipient (Brutus, Fleenor, and London, 1998). Self-assessments are also a common part of the process as these assessments provide a point of comparison with the other sources of feedback (Edwards and Ewen, 1996). The use and design of 360-degree programs varies by organization with some applying the process throughout the organization while others may only use it within a single department (London and Tornow, 1998). Most often, the process involves the various assessment groups completing survey questionnaires that provide feedback about the target individual. The surveys used for assessment may be internally generated questionnaires to address specific behaviors or competencies that the organization deems important. The surveys may also be standardized or customized assessments provided by outside organizations that address general leadership dimensions or managerial competencies (Lepsinger and Lucia, 1997).

Participation in a 360-degree program also varies with the needs of each organization. Many organizations reserve the process for upper- to middle-level managers and executives while others have implemented the program down to the level of individual contributors. Wide acceptance of 360-degree feedback within an organization is usually preceded by the acceptance of senior management; therefore most organizations begin the process at the senior management positions before administering to lower levels (Lepsinger and Lucia, 1997).

What a 360-degree program measures depends on the needs of each organization. Edwards and Ewen (1996) found that many organizations use 360-degree feedback to measure competencies that are relevant to the organization and that identify both high and low performance. Questionnaires usually contain items that assess a target manager's behaviors, skills, or perspectives (Van Velsor, 1998). Lepsinger and Lucia (1997) suggest that the program can be used to measure an individual's knowledge, skills, and style. Brutus et al. (1998) describe the program as one that measures individual items that may be grouped in broad performance dimensions such as administrative, communication, leadership, decision making, and personal motivation. Figure 1 further defines the knowledge, skills, and styles typically assessed by a 360-degree program as described by Lepsinger and Lucia (1997). Figure 2 lists the performance dimensions of Brutus et al. and indicates which rating sources are likely to observe those dimensions.

Figure 1.   Types of Data Collected by 360-degree Feedback
(After Lepsinger and Lucia, 1997)

Knowledge	Familiarity with a subject or discipline (e.g., knowledge of a business or industry)
Skill	Proficiency at performing a task; degree of mastery (e.g., ability to think strategically, communicate in writing, delegate work, influence, negotiate, operate a machine)
Style	Personal characteristics or ways of responding to the external environment (e.g., self-confidence, energy level, self-sufficiency, emotional stability)

Figure 2.   Performance Dimensions Likely to be Observed
By Different Rating Sources
(After Brutus et al. 1998)

Perfomance Dimensions	Subordinates	Peers	Supervisors
Administrative			X
Leadership	X		
Communication	X	X	
Interpersonal	X	X	
Decision Making		X	X
Technical		X	X
Personal Motivation		X	X

The presentation of feedback data to the target individual is equally important as collecting the data. Van Velsor (1998) suggests that the design of the report format can affect how easily a manager interprets the data and can also affect motivation to act on the feedback data. She found that most feedback reports use either graphic displays, narratives, or a combination of the two. Graphic displays present charts, tables, or graphs that show actual scores; and narratives provide descriptions and interpretations of the results. Regardless of how the data are presented, she states that most reports will provide a breakout of mean scores for each rating group on each item of the survey. Additionally, the recipient may be provided a comparison to normative scores of all individuals who have taken the survey to show where the target recipient stands in relation to colleagues, or he or she may be presented an "ideal" or "target" score that the organization has determined to be desirable for a particular item or area.

Once scores are tabulated and the report is prepared, organizations typically present the report to the target individual in one of three ways: one-on-one delivery, group workshops, or individual self-study (Lepsinger and Lucia, 1997). One-on-one delivery involves a coach or facilitator meeting individually with the recipient to assist with analysis and interpretation of the data as well as with the formulation of a personal development plan. Workshops provide data analysis, interpretation, and assistance with personal development plans to a group of individuals, usually ten to twenty, from the same level within the organization. The self-study method provides the recipient the feedback report and a self-paced guide, via a workbook or electronic program, to assist

with analysis, interpretation, and development plans. Each method has advantages and disadvantages. Lepsinger and Lucia (1997) note that one-on-one delivery usually provides the most interaction with the facilitator, a deeper explanation of individual results, and greater confidentiality of data as it is shared only with the facilitator. However one-on-one delivery requires considerably more time investment to complete the process than the other methods. Group workshops are more efficient than the other methods at providing similar information to a larger number of individuals. The group setting can also provide a more supportive environment for receiving negative feedback, especially when individuals see that they are not the only ones receiving negative feedback. Workshops can make the process more difficult for an individual who may need significant individual assistance in analyzing and interpreting feedback results. Self-study requires the least amount of time investment by the organization and provides the recipient with the greatest amount of confidentiality in personal data, but the lack of an individual or group facilitator means progress and development is largely dependent on the individual's motivation to act on the feedback data (Lepsinger and Lucia, 1997).

C. HISTORY OF 360-DEGREE FEEDBACK

Performance feedback has routinely been a part of the employer-employee relationship, yet this feedback normally was provided only by supervisors to subordinates. In the early 1950s the concept of management by objectives (MBO) emerged. Supervisors and subordinates worked together to identify objectives necessary to meet organizational goals and workers were provided more formal feedback targeted at their efforts toward achieving those objectives. Research found that employee productivity and job satisfaction improved when individuals were provided specific feedback on how well they met performance targets (Lepsinger and Lucia, 1997). As a result of this research, in the 1970s and 1980s companies began to use developmental feedback, in addition to performance appraisals and total quality management techniques, to improve individual and organizational performance (Edwards and Ewen, 1996).

In the 1990s many businesses began to adapt their organizational structure to meet the changing competitive environment by removing traditional hierarchical layers, increasing spans of control, and using self-directed teams (Edwards and Ewen, 1996).

These flatter organizations needed a more robust feedback mechanism than that provided by the standard supervisor-oriented feedback, and multi-source feedback began to fill this void.

Hedge, Boorman, and Birkeland (2001) offer a thorough review of the development of 360-degree feedback from the rating scale research of the early 1900s, through the beginning of upward feedback in the late 1950s, to the full implementation of multi-source feedback in the early 1990s. Two organizations that had the most influence in multi-source feedback development were the Center for Creative Leadership (CCL) and TEAMS, Inc. (Lepsinger and Lucia, 1997; Edward and Ewen, 1996). TEAMS, Inc. selected and registered "360° feedback" as a trademark for its proprietary multi-source feedback process in the 1980s. But it was *Wall Street Journal* reports in 1993 that brought the "360-degree feedback" label into the business press. When *Fortune* quoted General Electric CEO Jack Welch as saying he used 360-degree feedback, the practice attracted even greater attention and the term "360-degree feedback" became even more rooted as standard business vernacular (Edwards and Ewen, 1996).

D.   EMPIRICAL DATA ON 360-DEGREE PROGRAMS

While the increasingly competitive business environment was a factor in the development of 360-degree feedback, research that supported the effectiveness of this program as a development tool spurred the remarkable growth of acceptance and use within corporate America. Luthans and Peterson (2003) cite a recent survey that found nearly twenty percent of all American firms are using some type of 360-degree feedback program. The underlying theory of 360-degree feedback is that the ratings by different sources provide a target recipient with unique and meaningful feedback data on performance (LeBreton, et al., 2003). Most of the research of the 1990s supported this argument finding statistically significant differences across ratings provided by multiple sources. This research indicated that there was significant variation in ratings from supervisors, peers, and subordinates, and that this dissimilarity provided a feedback recipient with meaningful information from different perspectives within the

organization. Some recent research, however, questions the degree of uniqueness in multi-source ratings and also suggests that 360-degree programs may be less effective than originally believed.

1. Supportive Research

Support for the effectiveness of the 360-degree programs can be readily found in management, human resource, and psychological journals as well as the published works of subject matter experts of organizations in the leadership development industry. Brutus, Fleenor, and London (1998) argue that the multiple-rating sources are a main strength of 360-degree programs and that the multiple viewpoints have interesting differences. Based on their working experiences and the reviews of other studies, they conclude that feedback from multiple sources contributes to personal development and improved performance. Edwards and Ewen (1996) thoroughly discuss the potential of 360-degree feedback and suggest that outcomes can include improved employee satisfaction, behavior changes that are aligned with organizational objectives, and better team performance. They caution about the significant challenge of converting the potential of 360-degree feedback into a sustainable system; however they conclude that the program does have a measurable impact on the fairness of the assessment process, and that it is a useful development tool for an organization.

The study on upward feedback of student leaders and followers at the United States Naval Academy (USNA) is particularly pertinent to this thesis because of the military background of the participants (Atwater, Roush, and Fischthal, 1995). The subjects were 978 student leaders in their junior year and 1,232 student followers in their freshman year. The followers provided upward feedback to the leaders on performance in the area of general leadership behavior. The results suggested that leader behavior, as rated by followers, improved following upward feedback, and that leaders' self evaluations tended to become more similar to follower evaluations after feedback. Using a rating scale of one to five with five being the highest, mean follower rating scores improved from 3.77 to 3.99 and this improvement was significant at the one-percent level. The most notable improvements were seen in the leaders who initially rated themselves higher than they were rated by their followers.

10

Walker and Smither (1999) conducted a five-year study of upward feedback provided annually to 252 managers at a large, regional bank. The feedback survey was developed within the organization and was designed to assess behaviors believed to be associated with effective leadership, productivity, and implementation of strategic business objectives. The results showed that manager performance did improve and, similar to the USNA study, that the managers who initially received lower ratings from subordinates showed the most improvement. On a rating scale of one to five with one being the highest, mean feedback scores improved from 2.10 to 1.95 and this improvement was statistically significant at the one-percent level. Another finding from this study was that managers who held feedback discussion sessions with their direct reports improved more than mangers who did not conduct these sessions. This finding led the authors to assert that what a manager does with feedback affects the level of improvement generated by the feedback. A further indication from this study, based on its five-year run, was that improvements from upward feedback could be sustained over time.

Hazucha, Hezlett, and Schneider (1993) also conducted a study of 360-degree feedback effects over time. Their study involved managers who received feedback using an initial feedback report followed by another feedback report two years later. The feedback was provided via a Management Skills Profile (MSP) that measured managerial proficiency in various job-related dimensions such as administration, communication, cognitive and interpersonal skills, and overall leadership behavior. Their findings showed improved performance ratings at the second feedback opportunity and greater self-other rating agreement. On a rating scale of one to five with five being the highest, mean feedback scores improved from 3.66 to 3.74 and the improvement was statistically significant at the ten-percent level. Managers showing the most improvement were those who followed through on development with coaching and goal setting. The authors concluded that 360-degree feedback was an effective development tool.

Another longitudinal study on upward feedback produced similar results of effectiveness (Reilly, Smither, and Vasilopoulos, 1996). The study followed 92 managers who received four feedback surveys over a two and one-half year period. The surveys were designed specifically to measure behaviors in a supervisor-subordinate

11

relationship. Managers who initially received low to moderate feedback ratings showed the largest improvement at the second feedback administration six months later. Over the course of the entire study, the authors found that managers' improvements were independent of the number of times they received feedback, and that most of the performance improvement was observed between the first and second applications of the feedback. Using a rating scale of one to five with five being the highest, mean feedback scores improved from 3.75 to 3.92. Feedback scores for the lowest rated managers improved from 3.04 to 3.66. The mean improvement was statistically significant at the ten-percent level while the improvement for the lowest rated managers was significant at the one-percent level. The authors concluded that not only was the program effective, the improvement was not temporary and could be sustained over periods of time by periodically providing additional feedback.

The meta-analysis conducted by Kluger and DeNisi (1996) is an often cited work that both supports and contradicts the effectiveness of 360-degree feedback. Their work reviewed approximately 600 groups receiving feedback and the results showed that, on average, feedback could be associated with improved performance. The average effect, weighted by sample size, for all groups receiving feedback was 0.41 standard deviation units higher than groups not receiving feedback. This finding suggests that feedback has a moderately positive influence on performance. This finding is especially noteworthy because, unlike many studies that used only a pre-intervention and post-intervention comparison, Kluger and DeNisi compared groups receiving the intervention to groups not receiving the intervention. This comparison with control groups enables the results to be attributed directly to the intervention. Mitigating these results was the finding that, of those groups receiving feedback, about one-third showed improved performance, one-third showed little to no change, and one-third actually exhibited a decrease in their performance assessments. These findings appear to contradict the overall positive effect found for the entire study and may suggest that the 0.41 standard deviation unit improvement could have been caused by weighting the effects by sample size. Greater improvements may have been noted in larger group sizes and this would have introduced the positive skew in the overall results of the study.

Numerous other studies (Church and Bracken, 1997; Conway and Huffcutt, 1997; Greguras and Robie, 1998; Harris and Schaubroeck, 1998; Viswesvaran, Schmidt, and Ones, 2002) further support the effectiveness of 360-degree programs as performance development tools and the underlying theory of the unique and meaningful differences in ratings provided by multiple sources. These studies found that there is little similarity or correlation between the ratings assigned by different rating groups. Practitioners and researchers hold firm beliefs that multiple sources are superior to a single source when assessing behavior (Church and Bracken, 1997).

2.   Contradictory Research

More recent studies have introduced contradictory evidence on the theories and effectiveness of 360-degree feedback programs. While prior research had concluded that multiple-source ratings had meaningful differences because there is little correlation in ratings between sources, LeBreton et al. (2003) suggest these differences in ratings may be due to a statistical artifact that they describe as a restriction in variance in job performance. Their restriction in variance hypothesis is based on the assumptions that organizational interventions such as recruitment, selection, training, and counseling have been at least marginally effective, and that these interventions select and develop managers who then engage in relatively consistent behaviors across various situations and time. This restriction in variance in job performance, the authors argue, has caused past research to overstate the magnitude of the uniqueness in ratings from multiple sources.

The authors offer two competing hypotheses that may explain why previous research has concluded that multiple sources provide dissimilar ratings on the same target -- the discrepancy hypothesis and the restriction in variance hypothesis. They describe the discrepancy hypothesis as one that assumes raters from different sources observe different behaviors in a target manager, that managers behave differently around the different sources of raters, and that raters of different sources attach varying levels of importance to the same observed behavior in the target manager. Under this hypothesis, even though a manager may engage in relatively stable behaviors, raters from different sources have different perceptions of this behavior and thus assign different ratings.

When measured with traditional correlation-based indices, variation in ratings between sources has been determined to be statistically unique.

Under the restriction in variance hypothesis, LeBreton et al. (2003) argue that the distribution of managerial performance ratings is negatively skewed with the variance in ratings being restricted to the higher performance end of rating scales. They further argue that traditional correlation-based indices, such as Pearson correlations and intra-class correlations, are susceptible to downward bias when there is little between-target variance in ratings. In essence they are suggesting that different managers exhibit relatively little variance in overall performance, that this restricted variance in performance then restricts the variance in assigned ratings of that performance, and that this restricted variance in performance ratings causes traditional measures of correlation, used to measure the similarity between rating sources, to find little similarity between different sources of ratings. Because of the susceptibility of traditional correlation-based indices to downward bias when target behavior is restricted in range, the authors suggest that a new statistic, one that is unaffected by the restriction in variance in performance, should be used to measure correlations between different rating sources. They suggest the r_{WG} statistic, developed by James, Demaree, and Wolf (1984), as one that is unaffected by the restricted range in performance.

To test their hypothesis, LeBreton et al. (2003), conducted a Monte Carlo simulation and two large field studies of 360-degree programs. The Monte Carlo simulation involved the generation of 50,000 targets evaluated by four raters. The targets were then rank ordered according to their average ratings. After rank ordering, targets were gradually removed to simulate the recruiting, selection, and training interventions that would occur in a normal organizational setting. The simulation results showed that traditional correlation measures were downwardly biased when the range in performance was restricted while the r_{WG} measure was not affected by the range restriction. The Monte Carlo simulation confirmed their hypothesis that traditional measurements used in previous research likely overestimated the magnitude of differences in ratings between sources because their correlation indices were affected by restriction in variance. Their

14

independent field studies of 360-degree programs also showed that, under the restriction in variance hypothesis, different sources of ratings displayed significantly more similarity than previously estimated.

The conclusion of this study is that multiple sources of ratings tend to have substantially more agreement than previously believed, and that between-source rating agreement (e.g., peer-subordinate, supervisor-subordinate) is comparable to within-source rating agreement (e.g., peer-peer, subordinate-subordinate). This conclusion questions the belief in the superiority of multiple sources of ratings provided by 360-degree programs and questions whether the time and cost of administering these programs is greater than the potential psychometric benefits. The authors do suggest that, while the psychometric benefits may be marginal, there may still be psychosocial benefits gained from a 360-degree program such as increased job satisfaction, trust, perceptions of justice, and organizational commitment.

Another study looked at the effects of a rater's level in 360-degree ratings (Mount, Judge, Scullen, Sytsma, and Hezlett, 1998). Contrary to LeBreton et al. (2003), this study supports the theory of unique difference in ratings from multiple sources. However, the results of the study found that ratings by sources within the same level (e.g., two peers) were no more similar than ratings by sources from different levels (e.g., peer and subordinate). They suggest that rating differences among all raters are so unique that each rater should be viewed separately rather than aggregated by level. The authors argue that the current 360-degree practice of aggregating data by level is inappropriate and that this data averaging is mitigating valuable feedback information.

Scullen, Mount, and Goff (2000) studied the various factors that affect job performance ratings in a multi-source feedback setting. They developed a model that uses five factors they believe affect performance ratings in a multi-source assessment: ratee general job performance; ratee performance in a particular job dimension; rater idiosyncratic tendencies such as halo and leniency errors; rater organizational perspective (supervisor, peer, subordinate); and random measurement error. Using two data sets consisting of managers who received 360-degree ratings, the authors separated the variance in the ratings into three broad areas: the manager's actual job performance

(general and dimensional performance), rater bias (idiosyncratic effects and organizational perspective), and random measurement error. The authors used a correlated uniqueness-confirmatory factor analysis (CU-CFA) method to separate the rating variance of each rater into the three factors. The CU-CFA method is described as a two-step process where the CU method first divides observed variance into performance related and unique variance components. The second step uses CFA to divide the unique variance into rater-related variance and random measurement error. Scullen et al. determined that only approximately twenty-five percent of the variance in assessments could be attributed to a manager's actual performance while nearly fifty percent of the variance was due to rater bias effects. The authors concluded that, rather than being a true measure of manager performance, multi-source feedback largely measures the idiosyncrasies of individual raters. While this finding lends support to the underlying theory of using 360-degree feedback for developmental purposes, it suggests that multi-source feedback may introduce undesired bias in an administrative performance rating system.

Rather than examine rater effects on feedback, Greguras, Ford, and Brutus (2003) analyzed the level of attention that managers give to multi-source feedback ratings. An assumed benefit of 360-degree feedback is that multi-source ratings produce increased recipient self-awareness and improved performance (Mount et al., 1998). Greguras et al. (2003) suggest that an assumption of multi-source feedback programs is that recipients attend to the feedback information from each rating source. Their study was designed to test the hypothesis that feedback recipients attend to all sources of feedback in the same manner. They analyzed 213 managers in scenarios where multi-source ratings were varied across the different performance attributes of ability to lead others, administrative performance, building working relationships, and overall performance. The results indicated that feedback recipients did attend to all feedback ratings but not equally across all dimensions. Recipients attended to supervisor ratings more than peer ratings in all performance dimensions. Supervisor ratings were attended to more than subordinates' in all dimensions except building working relationships. Peer ratings were attended to more than subordinates' in the administrative performance dimension, and subordinate ratings were attended to more than peer ratings in the ability to lead others. This study supports

16

the theory that 360-degree feedback provides unique information from multiple sources and that recipients attend to the information from each source, but the results leave open the question of whether, as suggested by Figure 2, assessment tools should be tailored to the performance dimensions likely to be observed by particular rating groups.

Brett and Atwater (2001) tested the hypothesis that negative or discrepant feedback information motivates positive change in the recipient. Their study focused on recipient reactions to ratings and rating discrepancies across sources. The results indicated that less favorable feedback tended to produce negative feelings in the recipient and the belief that the feedback was less accurate. Further, if recipients viewed the feedback as less accurate, it was also viewed as less useful. Feedback that was viewed as less accurate and less useful did not consistently motivate positive change in the recipient. The meta-analysis of Kluger and DeNisi (1996) produced similar results when their analysis showed that feedback motivated positive change in only one-third of the recipients in the study.

Perhaps the most controversial finding links 360-degree feedback to a decrease in shareholder value (Pfau, Kay, Nowack, and Ghorpade, 2002). In their article the researchers discuss the Watson Wyatt 2001 Human Capital Index (HCI). This index is an ongoing study of how human capital practices relate to shareholder value in 750 publicly traded companies. The HCI scores were calculated in 1999 and again in 2001, and scores showed that companies using 360-degree feedback saw as much as a ten percent decrease in shareholder value. The controversy in this finding is whether shareholder value is a proper measure of human capital management effectiveness, especially in a time span of only three years (Chappelow, 2003). Chappleow argues that shareholder value is more often affected by other influences such as litigation, financial difficulties, and general market conditions. He cites work that suggests a better measure of the effects of human capital practices can be found in a combination of results such as revenues, earnings growth, and return on assets. Though the debate regarding this measure is certainly not resolved, the HCI findings suggest that organizations should thoroughly examine the expected costs and benefits of implementing a 360-degree feedback program.

17

London, Smither, and Adsit (1997) reviewed most of the pertinent literature on accountability in performance ratings and asserted that without accountability, 360-degree feedback would have little impact. Specifically they argue that raters should be held accountable for providing accurate feedback and that ratees should be held accountable for using the feedback. They also argue that the organization should be accountable for providing the resources to help support behavior change in feedback recipients. The researchers assert that, without accountability, 360-degree feedback can be inaccurate and easy to ignore. The authors concede that a dilemma exists between the accountability necessary for full realization of the benefits of 360-degree feedback and the expressed needs for anonymity of raters and confidentiality of the ratee's feedback. A psychologically-safe environment of anonymity and confidentiality is necessary to induce candid feedback, yet without accountability for accuracy and use, the program may be adding costs and limiting benefits.

E.   CONCLUSION

360-degree feedback is a development tool that presents a target recipient with performance assessments provided by self, supervisors, peers, and subordinates. The underlying theory of 360-degree feedback is that assessments from multiple sources provide unique and meaningful information to the recipient. The rapid growth in acceptance and use of 360-degree programs in corporate America was fueled by the need to adapt to the changing competitive environment and by numerous studies that supported the effectiveness of multi-source ratings. Although the majority of research supports the underlying theory of unique differences in multi-source ratings and the overall effectiveness of 360-degree feedback, recent research has raised questions about earlier findings and about the extent of benefits attributed to 360-degree feedback.

Results on the effectiveness of 360-degree programs are largely supportive but continued research is warranted. The current findings indicate that organizations should carefully consider the full range of expected costs and potential benefits when making decisions on implementing 360-degree programs for employee development.

III.   360-DEGREE FEEDBACK BEST PRACTICES AND LESSONS LEARNED

A.   INTRODUCTION

The phrase "360-degree feedback" is often used when describing organizational programs that use multi-source feedback surveys for personal development. For many organizations however, 360-degree feedback is only one part of a larger personal development program. This chapter examines studies of civilian organizations to identify best practices that enhance the benefits of using 360-degree feedback for personal development. A review of some current military 360-degree programs is also introduced to provide a more focused frame of reference for later comparison with the Surface Navy's 360-degree pilot program.

B.   CIVILIAN   BEST   PRACTICES   TO   IMPROVE   PROGRAM EFFECTIVENESS

1.   Executive Coaching and Feedback Workshops

The growth in popularity of executive coaching led Thach (2002) to study the quantitative impact on leadership effectiveness when using a 360-degree feedback process coupled with executive coaching. Her action research involved 281 executives and high-potential managers in a mid-sized, global telecommunications firm.   The organization used an external consulting firm to help customize a 360-degree survey to assess competencies necessary for leadership success within this organization. The main focus of the survey was to assess competencies deemed necessary to achieve the organization's five year business strategy. The study involved an initial 360-degree assessment followed by a training day that included an individual coaching session to debrief and analyze results. Members of the consulting firm served as executive coaches for the program and assisted the participants in preparing development plans to address no more than three areas identified for improvement and one area identified as a strength. Additional coaching sessions followed at one month, three months, and five months after

the initial session. The study concluded with the administration of mini 360-degree survey targeted at those areas identified for development during the initial coaching session.

The entire study was conducted in three separate phases. Phase one included development of the 360-degree survey and pilot testing the process on top executives including the CEO. Phase one data were not included in the program's analysis. Phases two and three were full implementations of the program. The second phase had 168 participants and the third phase had 113 participants. The participants in both phases completed a post-participation survey to provide their views on the program. The second and third phases were identical with the exception of minor modifications to the training day in the third phase that were suggested by participants in the second phase.

The results of the study indicated that leadership effectiveness ratings, as perceived by others in the mini-360 survey, had increased by fifty-five percent for the first group of participants and by sixty percent for the second group. The average number of coaching sessions completed, across both groups, was 3.6 as opposed to the four recommended by the program. While all participants who attended coaching sessions showed improved mini-360 self-scores in leadership effectiveness, Thach found that completing three to five coaching sessions had a much larger impact on improving self-scores than completing only one to two coaching sessions. Thematic analysis of the responses provided by participants through the post participation surveys revealed that thirty-four percent rated the coaching as the most positive part of the process and twenty-five percent rated the 360-degree feedback as helpful.

Thach cautions that her study is limited by its design as the analysis was of the complete process and could not accurately separate the effects of the coaching from those of the 360-degree feedback. An additional criticism is the lack of a control group to measure true program effect. Despite the limitations, this study suggests that 360-degree feedback coupled with executive coaching can have a positive impact on leadership development.

Luthans and Peterson (2003) conducted a similar study on the impacts of self-awareness coaching used in conjunction with a 360-degree feedback program. Their

study involved all employees, twenty managers and sixty-seven workers, of a small, Midwestern manufacturing company. As the entire organization was used in the study, supervisor, peer, and subordinate roles were all represented. The analysis focused specifically on the impact that the feedback and coaching combination had on manager self-awareness, which they defined as the difference between self-ratings and other's ratings, and on managers' and workers' attitudes. The authors developed a managerial feedback profile (MFP) to use for the 360-degree survey. The MFP assessed various behaviors in three broad areas: behavioral competence, interpersonal competence, and personal responsibility. Attitudes were assessed for all study participants through self-reports of job satisfaction, organizational commitment, and turnover intentions using other psychometrically accepted measurement instruments.

The study began with the initial administration of the MFP and attitude surveys. After completion of the surveys, the authors acted as feedback facilitators and coaches for the managers. The goals of the initial coaching session were to establish the manager's awareness of the discrepancy in self and other's ratings, to help managers determine why the ratings were different, and to help managers direct their increased self-awareness toward appropriate courses of action for improvement. No other coaching sessions were formally scheduled but the researchers did conduct random follow-up visits with each manager throughout the study period. The study was ended by re-administering the MFP and attitude measurement instruments to all participants three months after the initial assessment.

Study results showed that at initial assessment, manager's self-ratings were higher than other's ratings in all three factors. Scores on the follow-up MFP showed that the discrepancy between self and other's ratings had disappeared leading the authors to conclude that feedback and coaching positively affected the managers' self-awareness. Interestingly, the results also showed that the discrepancy reduction was not achieved by a lowering of self-ratings but by an increase in others' ratings of the managers. Attitudes of all participants also improved following the feedback and coaching. Participants reported increased job satisfaction and organizational commitment and decreased turnover intentions.

Luthans and Peterson acknowledge that the lack of a control group is a limitation in attributing results solely to the feedback and coaching. The design of the study did allow for measurement of change in attitudes but the absence of a control group prevents a clear determination that the improvements were caused directly by the feedback and coaching. The authors did not address any concerns with the relatively short period of the study. In view of the limitations, the authors suggest that 360-degree feedback with systematic coaching can have a positive effect on work attitudes and can possibly improve work performance.

Seifert, Yukl, and McDonald (2003) completed an analysis of feedback alone and feedback with coaching that used a control group to help assess actual program effects. The objectives of their research were to determine the effectiveness of a multi-source feedback workshop in changing managerial behavior and to determine if a skilled, neutral facilitator could enhance feedback effectiveness. Their study included twenty-one managers who received feedback from supervisors, peers, and subordinates. The managers were from two similar, regional savings banks. The managers were divided into three groups of seven. The experimental group received feedback via a facilitator led workshop, the comparison group received the same feedback reports but not in a workshop, and the control group received no feedback. The experimental and control groups were from the same bank while the comparison group was from the other bank.

The feedback instrument was developed to assess the influence behaviors of the managers. The feedback provided was a measure of the manager's use of influence tactics with others. The authors used previous research to identify four core tactics of managerial influence behavior: rational persuasion, inspirational appeals, consultation, and collaboration. A pre-measure survey was conducted for all twenty-one participants to provide a baseline assessment of the manager's use of influence tactics. A post-measure survey was completed three months later following the feedback intervention. The effect of the intervention was evaluated by measuring the change in a manager's use of influence tactics. Another survey was administered at the end of the workshop to assess manager's perceptions of feedback accuracy, feedback utility, and the capacity to improve based on feedback. The same survey was given to the comparison group with their feedback reports.

The feedback workshop was a seven-hour session held at the bank's training facility and the authors served as workshop facilitators. The facilitators first explained various tactics used to exert influence and showed a video demonstrating these tactics. Next the managers were given their feedback reports and facilitators offered advice on interpretation. The workshop then shifted to scenario exercises where the managers were presented a scenario and then worked in groups to develop an influence strategy for each scenario. The workshop concluded with facilitators assisting managers in developing action plans for using their feedback to improve influence behaviors.

The results of the feedback intervention showed that the experimental group significantly increased its use of two of the four core influence tactics, consultation and collaboration, while the control and comparison groups showed no significant change in any influence behaviors. The intervention evaluation surveys indicated that the experimental group and comparison group perceived no difference in feedback accuracy but the experimental group had a significantly higher perception of feedback utility and its capacity to improve performance. Based on the results the authors concluded that a feedback workshop can have a positive effect on changing behavior and that using a competent facilitator can increase the perceived utility of the feedback.

Rogers, Rogers, and Metlay (2002) conducted a survey of 145 global organizations that used 360-degree feedback. Companies such as Aetna, Allstate, Anheuser-Busch, Ford, Home Depot, Raytheon, and USX, were among the forty-three organizations that responded to the survey. The purpose of their survey was to determine how and why organizations are using 360-degree feedback. They divided the organizations into three groups, higher benefit, moderate benefit, and lower benefit, based on the organization's assessment of whether 360-degree feedback had been beneficial and if the 360-degree feedback process was worth the resources committed to the program. About twenty-one percent of the organizations considered 360-degree feedback to be of a high benefit, fifty-seven percent considered it of moderate benefit, and another twenty-one percent considered it to be of low benefit.

The survey results indicated that nearly ninety percent of the higher benefit organizations used coaching as part of their 360-degree feedback process. These

organizations reported investing significant time, resources, and control over the coaching process including selection and training of coaches. An interesting finding was that only twenty-five percent of the higher benefit companies used external coaches while fifty percent of the lower benefit companies used external coaches. The authors suggest this finding may be due to the expanded use of 360-degree feedback throughout the organization, which would make external coaching prohibitively expensive. Another possible explanation, though not suggested by the authors, is that internal coaches might have higher credibility with members of the organization than external coaches. The authors also state that, in a survey of 360-degree feedback participants, seventy percent reported that coaching helped them make better use of feedback results.

2.   **Anonymity and Confidentiality**

Confidentiality refers to the way in which a target manager's feedback data are shared, and anonymity refers to the protection of the identity of raters (Van Velsor, 1998). Absolute confidentiality and anonymity would be a situation where the feedback recipient is the only person who sees the data and the raters are completely unknown to the ratee. Van Velsor argues that confidentiality and anonymity are critical in the 360-degree process yet she concedes that limitations in the process preclude absolutes in either case. Edwards and Ewen (1996) also stress the need for both confidentiality and anonymity in the process. They recommend that feedback data be shared with a performance coach to enhance effectiveness but they caution against using the supervisor as the coach. Their argument is that the supervisor will face a dilemma of seeing feedback data that is to be used for development purposes only and then trying to forget these data when making performance appraisal decisions. A role conflict then occurs between the supervisor's position as coach for development and as judge for performance appraisal (Tornow, 1998). When confidentially barriers are broken in a developmental feedback process, feedback scores become less accurate and are usually inflated (Eichinger and Lombardo, 2003).

Eichinger and Lombardo (2003) cite recent surveys that showed half of supervisors in a 360-degree program had access to full feedback reports on their subordinates. They argue that this is a flawed practice rife with unintended consequences. They cite Antonioni's study (1994) that found non-anonymous direct

reports rated supervisors significantly higher than those whose ratings were anonymous as evidence of the problem with the practice. Citing their own studies, the authors found that average scores went up when raters were not anonymous and that forty-three of sixty-seven competency ratings increased significantly. Rogers et al. (2002) found that ninety-seven percent of the forty-three companies that responded reported that ensuring anonymity and confidentiality was a primary objective in their programs.

3. Training

Based on experiences with assisting in the implementation of 360-degree feedback programs, Edwards and Ewen (1996) argue that organizations that do not invest in training should not pursue 360-degree feedback. They suggest that training raters in how to properly provide feedback is equally important as training recipients in how to use the feedback. Rogers et al. (2002) found that companies reporting higher benefits from 360-degree programs were more likely to have invested in training for raters than lower benefit companies. Additionally, higher and moderate benefit companies were more likely to exert approval over the ratee's selection of raters than lower benefit companies. Ghorpade (2000) suggests that rater training should include detection of rater biases. This detection can be shown in trial rating sessions of hypothetical candidates who display wide variations in behavior. Raters are shown their own scores and the average of the group's of scores to reveal if they are habitually high or low graders. Ghorpade cites the work of Cascio (1997) as evidence that this "frame of reference" training can improve the accuracy of rater appraisals.

4. Use of Multiple Instruments

Martineau (1998) attempted to answer the question of how many times a particular instrument may be used for feedback. The heart of the question is whether a manager can learn anything new and meaningful from the same instrument used multiple times. She suggests that the flexibility of the instrument, such as the number of dimensions measured and variety of feedback provided, will determine how often it may be used. While offering no specific number, she does argue that saturation of any instrument for a particular individual will occur in time.

Using different instruments customized to the different ratee levels within an organization is another modification to the single instrument feedback program. Brutus

and Derayeh (2002), in their survey of Canadian organizations that use 360-degree feedback, found that approximately ten percent were using multiple instruments and these instruments were targeted to different segments within the organization. Rogers et al. (2002) found that higher benefit organizations used multiple instruments to measure the various sets of competencies expected at specific levels within the company. These organizations found that feedback targeted to specific job responsibility levels was more meaningful in employee development. Survey respondents reported that participants appreciated the targeted feedback instruments and that the customization helped individuals align their development goals with the larger goals of the organization.

5. 360-degree Feedback for Performance Appraisal

Dalton (1998) states that the practice of using 360-degree feedback for performance appraisal is controversial. She cautions against using 360-degree developmental feedback for appraisal because doing so violates the confidentiality of feedback data. She also suggests that use as a performance appraisal system ignores the research evidence that shows raters change their feedback scores if they are to be used for appraisal vice development only. Dalton does state that while some organizations have reported successful implementation of a 360-degree feedback performance appraisal system, a 1997 survey showed half of respondents that had used 360-degree feedback for appraisal had abandoned the practice for reasons such as negative employee reaction and inflated ratings. Scullen et. al. (2000) also urge caution as the results of their study suggest that, rather than measuring actual job performance, multi-source feedback systems largely measure the idiosyncrasies of the individual raters.

Ghorpade (2000) argues that the primary objective of 360-degree feedback is development rather than appraisal. He suggests that 360-degree programs should be used for development only but recognizes that, because of the costs of the program, many companies will desire to use them for appraisal purposes to increase return on investment in the program. In this instance, he suggests companies should use 360-degree feedback first as a development tool and only implement for appraisal after gaining wide acceptance within the organization. Lepsinger and Lucia (1998) also suggest a gradual approach. While leaning toward use for development only, they suggest that

organizations first begin with 360-degree feedback for development before proceeding to any use as a performance appraisal system.

Though they offer no empirical evidence, Eichinger and Lombardo (2003) suggest that use for performance appraisal can lead to rating coalitions where individuals agree to inflate each other's ratings as a form of protection from the threat of multi-source appraisal. Rogers et al. (2002) found that the process of moving from development to appraisal had often failed within the forty-three organizations that responded to their survey. They found that most organizations were using 360-degree feedback for development only and that higher benefit organizations were more likely to use 360-degree feedback only for development than were lower benefit organizations.

C. MILITARY PROGRAMS

1. Navy Flag/Senior Executive Service (SES) Program

Information on the Flag/SES program was obtained by personal communications with Mr. Jeff Munks (Jan, 2005) of the Executive Learning Office at the Naval Postgraduate School, and Dr. Roger Conway (Jan, 2005) of the Center for Creative Leadership (CCL) in San Diego, California. Additional information on the various survey instruments was obtained from the CCL website (CCL, 2005).

The Navy Flag/SES program is a joint effort between the Executive Learning Office at the Naval Postgraduate School and CCL. Newly selected Flag/SES personnel attend the Navy Flag Officer Training Symposium (NFOTS) as an orientation for their new positions. Prior to attending NFOTS, the participants are administered a battery of survey instruments, which include both 360-degree assessments and personality type indicators, to help each individual better understand self and to see how others assess their leadership competencies.

The two 360-degree assessments used are Benchmarks and the Campbell Leadership Index. Benchmarks is a CCL developed survey that assesses leadership skills, provides rater breakout and normative comparisons, and helps detect potential

flaws that could lead to career derailment. The Campbell Leadership Index provides the recipient with assessments of orientations toward leadership such as energy, affability, dependability, and resilience.

The personality indicators used include the California Psychological Inventory (CPI), the Change Style Indicator, the Myers Briggs Type Indicator, and the Fundamental Interpersonal Relations Orientation Behavior (FIRO-B). The CPI provides an assessment of personal and professional styles of interaction. The Change Style Indicator measures the individual's comfort level with change and approach to managing change. The Myers Briggs is the well known personality type indicator that measures four bipolar traits of personality: introvert-extrovert, sensing-intuition, thinking-feeling, and judging-perceiving. The FIRO-B instrument measures interpersonal effectiveness in the dimensions of inclusion, control, and affection.

During NFOTS the participants attend a coaching workshop where results of the various surveys are reviewed and interpreted. In addition to the coaching workshop, each participant meets one-on-one with an industrial psychologist for in-depth review of survey results and generation of personal development plans. After NFOTS, participants can request follow-on coaching sessions.

The combination of 360-degree assessments and personality type indicators provides participants with a well rounded view of self and with assessments by seniors, peers, and subordinates. The process is conducted only one time, during NFOTS attendance. The survey results are confidential, used only for personal development, and are not linked to any performance appraisal system. Based on feedback surveys, participants found the process to be beneficial and extremely valuable in helping them see self through the assessments of others.

2. Submarine Squadron Twenty

Submarine Squadron Twenty recently announced a 360-degree feedback pilot program, scheduled to begin in May of 2005, for the eight commanding officers in this unit (Spinner, 2005). The focus of the program is to provide participants a view of emotional and social leadership skills, to assess leadership competencies, and to highlight any behaviors that may be barriers to further advancement.

The Submarine Squadron Twenty program will consist of two survey instruments, a 360-degree feedback instrument and an emotional inventory instrument (Spinner, 2005). The program will use the LOMINGER VOICES Multi-rater 360 Assessment instrument and the BarOn Emotional Quotient Inventory. The 360-degree degree instrument will provide the recipient feedback data from supervisors, peers, and subordinates. The emotional inventory is a self-scored instrument and will complement the 360-degree assessment by providing the participant measures of competence in emotional and social functioning to better understand how decisions emotionally impact others.

The assessment program will consist of two formal sessions conducted on-site and one-on-one professional feedback tailored to each participant. Once feedback surveys are completed the participants will meet with an external executive coach to interpret the results. Following the individual sessions the commanding officers will participate in a group session to debrief results and develop improvement goals based on their results. Each participant will also receive a developmental coaching guide and a telephone follow-up interview with their executive coach.

D.   CONCLUSION

Civilian organizations have adopted additional practices to enhance the benefits of their 360-degree assessment programs. One of the most beneficial practices identified is using a coach or feedback workshop to assist with the presentation and interpretation of results and the formation of personal development plans. Higher benefits are also achieved when 360-degree assessments are used for development and not appraisal purposes, when raters are trained in how to provide proper feedback, and when multiple instruments are used to target competency development at specific levels within the organization.

The limited numbers of existing military programs have incorporated many of these best practices into their processes. They invest heavily in professional coaching, use personality indicator instruments in addition to 360-degree assessments to provide a more robust view of self, and use the entire process for development purposes only.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. NAVY 360-DEGREE FEEDBACK PILOT PROGRAM

A. INTRODUCTION

The Navy's formal performance appraisal system provides only top-down feedback from one constituent, the reporting senior. Additionally, the Navy-wide leadership development program provides leadership training in formal classroom settings and electronically via electronic learning resources. Despite broad acceptance within corporate America, the Navy currently has not institutionalized a service-wide multi-rater leadership development program.

This chapter presents a description of the current appraisal and development process, provides a detailed description of the Surface Warfare community's 360-degree feedback pilot program, and presents a comparative analysis of the pilot program with identified research evidence.

B. WHY 360-DEGREE FEEDBACK?

1. Current Appraisal and Development Process

The Navy's current performance appraisal process is the Fitness Report (FITREP) and Evaluation (EVAL) program delineated in the Naval Personnel Command instruction BUPERSINST 1610.10 (1995). FITREPs are provided to senior enlisted and officer personnel and EVALs are provided to junior enlisted personnel. This program provides top-down feedback from one reporting senior who rates the individual's past performance in areas such as professional expertise, military bearing, mission accomplishment, and leadership. Reports are produced and presented to each individual annually. Six months prior to the formal report, each member receives a one-on-one, mid-term counseling session with his or her reporting senior to discuss previous performance and to address any areas that may need performance improvement before the formal report is written.

The Naval Personnel Development Command (NPDC) has primary responsibility for personal and professional development within the Navy (NPDC, 2005). The Center for Naval Leadership (CNL), a subordinate command of NPDC, operates over twenty learning sites at most major naval installations within the United States and overseas.

CNL provides leadership development training through courses taught at the learning sites and by mobile training teams (MTT) when there is a need at a location without an established learning site. The courses range from first-line leadership development, targeted to the most junior leaders in the Navy, to the advanced officer leadership course for senior Navy leadership. The courses last approximately two weeks and cover leadership skills and competencies necessary for the respective leadership positions. The Navy's goal is to have each individual complete the appropriate leadership development course before assignment to a leadership position (Naval Administrative Message [NAVADMIN], 2004).

In addition to formal classroom instruction, NPDC also developed Navy Knowledge Online (NKO), a web portal designed as an electronic delivery vehicle for NPDC products. Through NKO, Sailors may access various courses on leadership, professional performance, and personal development. NPDC describes NKO as a single point where any Sailor may access information on career issues (NPDC, 2005).

2. Supplementing Current Appraisal and Development Processes

The widespread popularity of 360-degree feedback as a management development tool in corporate America led the Navy to institute a similar program for its most senior leaders, the flag officers. The success of the flag officer program over the past four years and the lack of a Navy-wide, multi-rater leadership feedback program have provided further impetus for the Navy to institute a service-wide 360-degree program for leadership development.

In July of 2004 the Surface Warfare Commanders Conference recommended that the Surface Warfare Officer (SWO) community be used as a test group for a 360-degree feedback pilot program. Results of this pilot program will be used to assess the feasibility of implementing a Navy-wide 360-degree feedback program.

C. 360-DEGREE FEEDBACK PILOT PROGRAM DESIGN

All of the following information on the 360-degree pilot program was obtained from the NKO 360-degree resources web page and by personal communications with LCDR Jim Pfautz (Jan-Apr, 2005), the 360 Project Lead at CNL.

1. Pilot Phases and Participating Units

The pilot program will be administered in three separate phases over a three-year period. Phase 1 began in October, 2004 and ended in November, 2004. Phase 1 was not a full implementation of the pilot as only six ships and one shore command participated. Phase 1 was not designed to collect data for statistical analysis but rather to identify any obstacles with the software program and internet connectivity.

Phase 2 is a full implementation of the pilot program. This phase began in January, 2005 and is scheduled to continue until October, 2006. Approximately 450 personnel from sixteen ships and three shore commands (see Figure 3) will participate in this phase. Individuals receiving 360-degree feedback assessments will include Surface Warfare Officers and Supply Corps Officers in the grades of Ensign (O-1) through Commander (O-5), the Command Master Chief Petty Officer (E-9), and other Master Chief Petty Officers (E-9) assigned to the Phase 2 participating commands.

Figure 3.   Phase 2 Participating Ships and Shore Commands

USS LAKE CHAMPLAIN (CG-57)	USS VELLA GULF (CG-72)
USS PRINCETON (CG-59)	USS LEYTE GULF (CG-55)
USS JOHN PAUL JONES (DDG-53)	USS MITSCHER (DDG-57)
USS PINCKNEY (DDG-91)	USS DONALD COOK (DDG-75)
USS MCCLUSKY (FFG-41)	USS CARR (FFG-52)
USS JARRETT (FFG-36)	USS NASHVILLE (LPD-13)
USS CLEVELAND (LPD-7)	USS WHIDBEY ISLAND (LSD-41)
USS GERMANTOWN (LSD-42)	USS CARTER HALL (LSD-50)
Surface Warfare Officers School	Surface Warfare Development Group
Afloat Training Group Pacific	

Phase 3 is scheduled to begin in October, 2006 and to continue until September, 2007. Phase 3 will be similar to Phase 2 with approximately the same number of ships and shore commands participating, although specific ships and shore commands have not yet been designated. The results of Phase 2 will be used to inform decisions about any changes or improvements to Phase 3; therefore the specific design of Phase 3 is yet to be determined.

2.   Survey Instrument

The pilot will use a single instrument in Phase 2 for all participants. The survey instrument, created by CNL, is a web-based, customized 360-degree feedback survey designed to assess individuals in the five core areas of the Navy Leadership Competency Model: accomplishing mission, leading people, leading change, working with people, and resource stewardship. These five core competencies are divided into twenty-five sub-competencies. Figure 4 lists the Navy's five core leadership competencies and their associated sub-competencies.

Figure 4.   Navy Core Leadership Competencies and Associated Sub-Competencies

Accomplishing Mission	Leading People	Leading Change	Working with People	Resource Stewardship
Responsibility, accountability, and authority. Decisiveness/risk management. Continuous improvement. Problem solving. Technical credibility.	Developing people. Team building. Combat/crisis leadership. Conflict management. Leveraging diversity. Professionalism.	Creativity and innovation. Vision. Strategic thinking. External awareness. Flexibility. Service motivation.	Influencing and negotiating. Partnering. Political awareness. Oral communications. Written communications.	Financial management. Leveraging technology. Human resource management.

The survey contains sixty-eight specific questions to assess the twenty-five sub-competencies. For most of the core competencies, two to three questions are used to assess each of the sub-competencies. However, in the leading change core competency, only seven survey questions are used to assess the six sub-competencies.

Each of the survey questions will be answered using an "extent-based" scale with a scale range of one to five. For each question the rater will assess how often the target individual accomplishes that task or displays that behavior. A response of one indicates "never"; two indicates "some extent"; three indicates "slight extent"; four indicates "great extent"; and five indicates "very great extent." Appendix A lists each of the survey questions and associated core leadership competencies.

3. Feedback Reports and Development Plans

Individual feedback reports are generated after all surveys are collected, aggregated, and validated by the feedback software program. Once the survey process is complete, members may access their feedback report via the 360-degree program website. The feedback report displays the target individual's scores in each of the twenty-five competency areas. Scores are broken out by each rating group (supervisor, peer, subordinate, and self), and an overall mean score of all responses, including self, is computed for each sub-competency. Additionally, a normative score is computed for each competency. The normative score for each competency is the average score that each rank (e.g., LT, LCDR) has received from all ratings groups based on all survey responses to date. If the target individual's mean score is lower than the normative score, that competency is identified as an actionable development opportunity. If the individual score is higher than the normative score, no improvements are indicated as necessary for that competency. For example, a lieutenant might receive a feedback report with a mean survey score (average of supervisor, peer, subordinate, and self) of 3.5 in the financial management competency. The financial management normative score for a lieutenant (based on the average of all surveys from all rating groups to date) might be 4.0. The financial management competency would then be identified as a development opportunity.

An Individual Development Plan (IDP) is also generated by the 360-degree program. The IDP lists all the competencies identified as development opportunities and

provides a development guide to address those deficiencies. Included in the IDP is an embedded link to the IDP Resources page hosted at NKO. The NKO web portal has a resource page for each major competency area. The resource page for each competency area has links to various on-line training aids and electronic learning courses to assist in development of those sub-competencies identified as deficient.

Of the competencies listed in the IDP as development opportunities, the feedback recipient will identify those competencies that he or she feels are most in need of improvement. While many competencies might be identified as development opportunities, the individual will select a small number, approximately two to four, to target for development during that assessment period. Using the IDP as a guide, the recipient will develop an action plan to address those two to four competencies deemed most in need of improvement. While there is no standard format for an action plan, the plan is based primarily on the deficiencies highlighted in the IDP and the NKO training resources identified as measures to assist in improving those deficiencies. The IDP and action plan will be discussed with the Commanding Officer at the mid-term counseling. It should be noted that the action plan developed in the pilot program is largely a training plan that uses NKO resources to develop deficiencies, whereas most development plans in the literature, thought not discussed in detail, appeared to use a more "whole person" developmental approach and included items such as behavioral objectives in addition to deficiency improvements.

4.   Business Rules for Pilot Administration

The 360-degree program website and software program that manages the feedback survey administration and compilation processes is operated by an external contractor. ALUTIIQ was awarded the management contract for Phase 2. Participating commands and CNL jointly manage program participation. CNL provides initial program training and the commands select participants and manage the program.

Each participating command will select a command member to serve as the focal point for the program. This individual will be selected based on familiarity with the command and command members, and will be responsible for administration of the program within that command. The command focal point will also be responsible for selecting raters for the feedback recipients.

All command members, E-9 through O-5, who have been at their command for a minimum of 120 days, will participate in the program. Each member will receive an initial 360-degree assessment approximately one month prior to his or her FITREP mid-term counseling session. The timing of the initial assessment allows for collection of all feedback surveys, for generation of the feedback report, and for generation of the Individual Development Plan (IDP). The individual's feedback report is confidential and will not be seen by the Commanding Officer. The recipient will forward the IDP to the Commanding Officer for review prior to the mid-term counseling session. The member will bring the action plan to the mid-term counseling and will discuss both the IDP and action plan with the Commanding Officer. The Commanding Officer will be able to assess the individual's action plan, determine if the action plan is appropriate based on the development opportunities listed in the IDP, and recommend changes to the action plan if necessary.

A second 360-degree assessment will be administered six months following the first assessment. This assessment will be identical to the first with both a feedback report and IDP generated by the program and a member-developed action plan to address the deficiencies noted in the IDP. The second assessment will enable measurement of development progress since the first assessment. As the second assessment will occur one month prior to the formal FITREP, the IDP generated during the second assessment will be shared with a mentor, but not with the Commanding Officer, to prevent any association of the developmental feedback with the FITREP performance appraisal. There are no formal guidelines for the mentor process, however the mentor will most likely be selected by the individual and may or may not be involved in the first 360-degree assessment process.

D.   PILOT PROGRAM ANALYSIS

1.   The Survey Instrument

The survey instrument appears to be properly aligned with the Navy's strategic vision of successful leadership traits in that it seeks to measure specific behaviors that support the Navy's five core leadership competencies. However, the psychometric

validity of the instrument can not be determined by this thesis. As the Navy's leadership competencies apply to all ranks of Navy leaders, the instrument used is the same for all participants.

The use of a single instrument for all participants can have disadvantages. Parts of the instrument may not be able to accurately assess each leadership competency across all ranks. For example, the most junior officers may have little or no involvement in budgeting or resource allocation decisions because of their position within the command. Raters may not be able to give ratings in these areas, or when given, the ratings may be inaccurate or not applicable. Instruments modified to target specific behaviors expected to be mastered by different levels of responsibility may be more beneficial than a single instrument measuring each area equally across all levels in the command. The use of multiple instruments can present the recipient with new developmental feedback during regular career progression. Research has shown that organizations report higher program benefits when using multiple instruments targeted to specific levels of responsibility rather than using one instrument across all levels of responsibility (Rogers et al., 2002).

Research evidence also suggests that recipients do not attend equally to all sources of feedback across all competency areas. Gregarus et al. (2003) found that while recipients attend to supervisor ratings more than others, they attend to subordinate ratings more than peers, in the ability to lead others and to peers more than subordinates in general administrative areas. The single instrument may be presenting the recipient more feedback than he or she will actually use. Instruments that can be modified to provide feedback from sources that the recipient will actually attend to, such as leadership feedback only from supervisors and subordinates, may be more beneficial than an instrument that provides feedback from all sources across all measured dimensions.

The use of a single instrument over time can also increase the potential for saturation. As an example, an Ensign (O-1) who remains in the Navy and is regularly promoted, can expect to achieve the rank of Lieutenant Commander (O-4) in approximately ten to eleven years. Over the course of his or her career, this person would have received twenty or more applications of the same instrument. One can reasonably assume that the instrument will have lost its developmental impact for this individual.

Research has shown that most improvement occurs between the first and second application of an instrument and that this improvement can be sustained over time with occasional re-application of the instrument (Reilly et al., 1996; Walker and Smither, 1999). Less frequent application of a single instrument may lengthen the time that the instrument remains viable as a development tool. Additionally, the use of instruments tailored to the various levels in the organization, as described above, would present the recipient with varied instruments through career progression and may also, therefore, reduce the problem of saturation.

2. The Feedback Report and Development Plan

The feedback reports present the recipient with scores broken out by rating group and with normative scores to use for comparison. The breakout of group scores, averaging of scores across all groups, and use of normative scores for comparison are common practices in many 360-degree programs. In the pilot program, including self-scores in the average of all group scores may contaminate the process of identifying competency areas for development. The overall mean rating, which includes the self-score, is used to compare to the normative score for each assessed area. If the mean score in a specific area is above the normative score, that area is not identified as a development opportunity. Previous research studies found that self-scores often differed, sometimes significantly, from other groups' ratings (Atwater et al., 1995; Hazucha et al., 1993; Luthans and Peterson, 2003). Additionally, more improvement was seen in individuals who initially had higher self-ratings than others' ratings. Including the self-rating score in the mean rating score can potentially distort this score and thus affect the normative comparison. If a self-rating is significantly lower than other ratings, the mean score would be averaged downward and this competency area could incorrectly be designated as one that needs improvement. Conversely, a significantly higher self rating could increase the mean score rating and could incorrectly identify a competency as an area where no improvements are needed.

The presentation of results through the IDP, the development of a Commanding Officer- or mentor-approved action plan, and the use of individual electronic training resources, is a development method that most closely resembles a self-study process. Self study is one of three ways that most organizations provide feedback analysis to the

recipient, the other two being through an individual coach or through facilitator-led workshops (Lepsinger and Lucia, 1997). While research has shown that executive coaching coupled with multi-source feedback has a significantly positive impact on development and improvement (Thach, 2002; Luthans and Peterson, 2003; Seifert et al., 2003), this process is also the most costly and time consuming. For the pilot program, and for any future Navy-wide program, executive coaching for each participant would almost certainly be prohibitively expensive. The pilot program self-study method, linked to specific training aids and courses at NKO, provides a cost-effective method of delivering developmental assistance to a large number of participants. However, more elaborate self-directed action plans, which include behavioral objectives as well as deficiency improvements, may provide greater value for both the individual and organization than do plans that rely only on NKO training resources.

3. The Process

The pilot program is specifically intended to be used for development purposes only and this type of use is consistent with research evidence. Organizations receiving the most benefit from a 360-degree program reported using the program for development purposes only (Rogers et al., 2002). Most experts support the idea that the program is better suited to development rather than appraisal (Dalton, 1998; Lepsinger and Lucia, 1997). Feedback recipients only share their IDP and action plan, not feedback report scores, with their Commanding Officer, and these are shared with the Commanding Officer only during the mid-term counseling session. The IDP and action plan developed in the assessment prior to the formal FITREP are not shared with the Commanding Officer but with a mentor. While this process is a positive step in ensuring that feedback remains developmental and is not linked to the performance appraisal process, it raises a question about why this assessment occurs. An annual administration of the survey during the mid-term FITREP cycle could also reduce the risk of entangling developmental feedback with the performance appraisal process and could reduce the potential rate of instrument saturation.

The pilot program will use a command focal point for local administration of the program to include selection of raters. Selection of raters by someone other than the feedback recipient increases the level of anonymity of raters, which is necessary to ensure

raters provide honest feedback without fear of reprisal. Rater selection by the command focal point can ensure that more accurate feedback is provided because raters are selected based their familiarity with the target individual. Survey research has shown that organizations reporting moderate to high benefits from 360-degree feedback were much more likely to have an administrative approval process for the selection of raters than those organizations reporting lower benefits from 360-degree feedback (Rogers et al. 2002).

Ratee accountability in the pilot program is enhanced by the process of sharing the IDP and action plan with the Commanding Officer and other mentors. Experts argue that without accountability for action, target recipients may do nothing with their feedback, thus the program would provide little benefit to the organization (London et al., 1997). Commanding Officers can compare the individual's action plan to the IDP generated by the survey program and offer advice for improving the action plan if necessary. Sharing the follow-up assessment IDP and action plan with a mentor allows the mentor to determine what, if any, developmental progress has been achieved and whether or not the individual completed the action plan created during the previous assessment. In this process, the Commanding Officer and mentor provide an accountability mechanism and supplement the program's self-study method of development by acting as internal coaches for the target individual. Internal coaches were more likely to be used by organizations reporting higher benefits from 360-degree feedback (Rogers et al., 2002).

E.   CONCLUSION

The Navy's current processes for performance appraisal and personal leadership development are the formal FITREP and EVAL program and the CNL leadership development courses. These processes provide valuable performance feedback and leadership training information to each individual; however they lack the multi-source-perception feedback of a 360-degree program. The popularity of 360-degree feedback in corporate America and the success of the Navy Flag/SES 360-degree program have induced the Navy to analyze the feasibility of introducing a Navy-wide 360-degree feedback program.

The Surface Warfare community is conducting a three-year trial of a 360-degree feedback program to provide data for analysis of potential Navy-wide implementation. While many aspects of the program appear to be largely in line with previous research evidence and with identified best practices, others are not. The use of a frequently applied, single survey instrument, a narrowly focused individual action plan, and the inclusion of self-scores in the average presented on the feedback report are not in accordance with the literature or best practices; therefore suggested improvements include adjustments to the survey instrument and feedback reports and the use of more broadly focused action plans.

V. PROGRAM EVALUATION

A. INTRODUCTION

Evaluation is essential to determine the effects of any program that is introduced to accomplish some goal or effect some change. Proper evaluation design is necessary to enable evaluators to determine the gross effects of a program and to be able to separate the net effects attributable to the program from the gross effects. While evaluation should be a part of every program implementation, many organizations do not expend the effort to formally evaluate programs, especially 360-degree programs. Rogers et al. (2002) found that, of the companies that reported receiving high benefits from 360-degree feedback, over fifty-five percent evaluated their programs. Of those companies that reported receiving low benefits from 360-degree feedback, only thirty-five percent performed evaluations.

This chapter introduces general and specific concepts in program evaluation. These evaluation concepts are then applied to the Surface Navy's 360-degree pilot program to develop a proposed evaluation plan for use when pilot program data become available.

B. HOW TO EVALUATE A PROGRAM

1. Use of Evaluation Findings

Patton (1997) suggests that evaluation findings generally serve three purposes: making judgments, identifying improvements, and producing knowledge. Judgment-oriented evaluations are most often used to assess whether or not a program actually works. Improvement-oriented evaluations may be used to identify areas of a program that need adjustment. Knowledge-oriented evaluations are largely conceptual and influence thinking or build theory about a specific program or concept, e.g., building theory about whether there is a superior method of training delivery. Judgment- and improvement-type evaluations most often induce a decision or some type of action on a program while knowledge evaluations do not necessarily induce decisions but rather help to generate a better understanding of the program being evaluated.

Patton does state that all three processes support decision making but that the decisions based on each process can be different. Judgment evaluations are used to determine the overall merit or value of a program and whether or not that program should be continued.   Improvement evaluations support decisions about how to make adjustments to ongoing programs. Knowledge evaluations typically inform decisions about larger policy issues. Figure 5 lists some specific examples of uses for each type of evaluation.

Figure 5.   Primary Uses of Evaluation Findings
(After Patton, 1997)

Evaluation use	Examples
Judgment	Summative evaluation
	Accountability
	Cost-benefit decisions
	Decide a program's future
Improvement	Formative evaluation
	Identify strengths and weaknesses
	Continuous improvement
	Manage more effectively
Knowledge	Generalizations about effectiveness
	Extrapolate principles about what works
	Theory building
	Policy making

2.   Impact Assessment

Rossi and Freeman (1989) state that impact assessments are used to determine whether or not a particular program or intervention produces the intended effects. The aim of impact assessment is to produce an estimate of the net effects of the particular program to provide data to support decisions about the program. To estimate net effects, an evaluation must be able to separate the effects caused by the intervention from those caused by other influences. The methods used to measure program effects usually fall

into one of two categories: experimental or quasi-experimental designs, and non-experimental designs (Posavac and Carey, 1989).

Experimental and quasi-experimental designs normally involve participants sorted into two or more groups. One group is designated as the control group and does not receive the intervention or participate in the program, while the experimental group or groups undergo the intervention or participate in the program. Measurements are normally taken prior to and following the intervention for both groups and differences are attributed to the program or intervention (Rossi and Freeman, 1989).

True experimental designs randomly assign participants to both groups, whereas quasi-experimental designs have participants that self-select or are selected by administrators for participation. Because quasi-experiments use participants not selected at random, various experimental designs are available. The most frequently used quasi-experimental design is the matched control group where program administrators select control group participants that most closely resemble the characteristics of those in the experimental group (Rossi and Freeman, 1989).

Non-experimental design typically involves only the experimental group in the analysis of program effects. Measurements may be taken on the experimental group following the intervention, a posttest design, or they may be taken before and after the intervention, a pretest-posttest design. (Posavac and Carey, 1989). Other non-experimental impact assessment methods include time-series analysis, where repeated measurements are taken on the experimental group over an extended period of time, and subjective judgments of effectiveness by the program administrators and participants, which are usually gathered by surveys (Rossi and Freeman, 1989).

Impact assessments that provide the most accurate measurement of program net effects are those of the experimental and quasi-experimental design (Rossi and Freeman, 1987; Posavac and Carey, 1987). The use of control groups in experimental and quasi-experimental designs provides greater validity than non-experimental designs in determining effects that are attributable to the program under study. Rossi and Freeman (1987) also argue that experimental and quasi-experimental designs are more appropriate than non-experimental designs in studying partial-coverage programs, i.e., programs

where only a portion of group members receive the intervention, as there are participants readily available to use in control groups. They further assert that the decision to assess by experimental or non-experimental design should be based most heavily on whether the intervention is a full-coverage or partial-coverage program. A disadvantage of experimental and quasi-experimental designs is that they are more difficult to construct and are usually more costly and time consuming than non-experimental designs.

Non-experimental designs are less accurate than experimental designs in measuring a program's net effects and are most often used in full-coverage programs as there are no members available to use as controls. The weakness of non-experimental designs is that they capture effects that can be attributed to sources other than the intervention such as participant maturation and experiences outside the program (Rossi and Freeman, 1989). The most frequently used non-experimental design is the pretest-posttest design, which is often referred to as before-and-after studies. This type of assessment simply measures participants before the intervention and after the intervention to determine program effects. While this type of design does allow some inference about whether program effects are positive or negative, the magnitude of the effects attributable to the program can not be determined. Despite this drawback, pretest-posttest designs do present information about program impact and can serve as the basis for more in-depth analysis through experimental or quasi-experimental design (Rossi and Freeman, 1989). Time-series analysis can improve assessment of actual program effects as participants are measured repeatedly over time, but in most social intervention programs time-series analysis normally must continue for a period of years to yield results. Subjective judgments by administrators and participants are the least accurate for determining program effects but they may contribute valuable information about program operation that can lead to refinements in the program to increase satisfaction or participation (Rossi and Freeman, 1987).

3. Implementation Analysis

Patton (1997) describes implementation analysis as an evaluation to determine if all the parts of a program are working correctly and if the program as a whole is working as it was intended. He suggests that while assessing program outcomes is important, equally important is understanding what happened in the program that can reasonably

account for the outcomes. Patton asserts that improper implementation can lead to erroneous decisions to terminate or expand a program. He offers variations that can be used individually or in combination to evaluate implementation: effort evaluation; process evaluation; component evaluation; and treatment specification.

Effort evaluation focuses on the activities that take place within the program and assesses the level of input from participants and administrators. This type of evaluation seeks to determine participation levels and completion rates of a program and whether or not administrators provide all necessary resources for proper functioning of the program. Process evaluation focuses on the operations of the program to determine strengths and weaknesses. Process evaluation looks at how the outcomes are produced and seeks to explain successes, failures, and changes in a program. Items in a process evaluation may include participant and administrator perceptions of the program as well as investigations of informal or unintended processes that develop within the program. Component evaluation assesses the distinct parts of a program to determine how they are working within the larger program system. Finally, treatment specification involves measuring the intended effect of the program. Treatment specification identifies the independent variables believed to affect outcomes, measures the outcomes, and attempts to determine if the treatment causes the outcomes (Patton, 1987). Patton's treatment specification is comparable to the impact assessment of Rossi and Freeman (1989) in that it attempts to determine causality, however in implementation analysis, treatment specification also attempts to determine if treatments are administered equally across all groups and if knowledge can be gained about the treatments that may influence policy or decisions elsewhere. Figure 6 lists some possible questions that may be used in implementation evaluations.

Figure 6.   Sample Implementation Evaluation Questions
(After Patton, 1987)

Effort Evaluation

- What do participants actually do in the program?

- What are the participant's primary activities and experiences?

Process Evaluation

- What are the programs key characteristics as perceived by various stakeholders? Are these perceptions similar or different? What is the basis for difference?

- What do the participants like and dislike?

- What has changed from the original design and why/

- What has been learned that might inform similar efforts elsewhere?

Component Evaluation

- What's working as expected? What's not working as expected?

- What are the participant's perceptions of what is working and not working?

Treatment Specification

- Can the program be modeled as an intervention or treatment with clear connections between inputs, activities, and outcomes?

- What assumptions have proved true?

- What aspects are likely situational and what aspects are likely generalizable?

---

4.   Efficiency Analysis

Efficiency analyses provide a framework for administrators to evaluate a program's outcomes in relation to the program's costs. Cost-benefit analysis compares costs to outcomes and both are estimated in monetary terms. Cost-effectiveness analysis is used when benefits can not be quantified in monetary terms and compares program outcome units to monetary costs (Rossi and Freeman, 1989). Posavac and Carey (1989) argue that outcomes of programs can not be fully evaluated unless their costs are considered in the evaluation. Cost analyses are used to make judgments about the value of program outcomes, to make decisions about whether or not to continue a program, and to make comparisons of multiple programs to determine which provides the greatest benefits with the least costs.

Cost-benefit analysis is conducted by calculating all costs associated with a particular program. Depending on the characteristics of the program, costs may be grouped into a variety of categories: fixed and variable, sunk and incremental, recurring and non-recurring, direct and indirect (Posavac and Carey, 1989). Regardless of the nature, all costs attributable to the program must be included to conduct a cost-benefit analysis. Benefits of the program are quantified in the same monetary units as the costs and are then compared to the costs. If benefits exceed costs the program produces net benefits. Conversely, if costs exceed benefits the program produces net costs. Program administrators must then determine if the benefits of a program are sufficient to justify the costs of providing those benefits (Rossi and Freeman, 1989).

Cost-effectiveness analysis is conducted similarly to cost-benefit analysis except that benefits are not quantified in monetary units. All costs attributed to the program are calculated and measured against the outcome units of a particular program. An example is a program designed to improve student standardized test scores. Test score improvement can not be easily quantified in monetary terms so the score improvement is used as a measure of effectiveness. The program is evaluated on the costs necessary to achieve improved scores. Cost-effectiveness analysis is especially useful in comparing programs designed to produce similar results, such as improving test scores. Programs can be measured and rank ordered based on costs to produce a specific level of score improvement or based on the magnitude of improve per unit of cost (Rossi and Freeman, 1989).

One cost that is often overlooked and also very difficult to quantify is opportunity cost (Rossi and Freeman, 1989; Posavac and Carey, 1989). Opportunity costs occur due to the nature of limited resources and are reflected in the costs of selecting one alternative over others. An example is the decision to attend college full time. A student who decides to attend college gives up the opportunity to work full-time. The costs of not working are the opportunity costs in this decision. In many organizational human resource programs, the participant's time is the greatest opportunity cost. The time necessary to participate in a program is time that could instead have been spent performing work for the organization (Posavac and Carey, 1989). Opportunity costs

often can only be estimated based on assumption and thus they may be quite controversial in any efficiency analysis (Rossi and Freeman, 1989).

A balanced scorecard approach may also be used to assess the effectiveness of a program. The balanced scorecard is a strategic management process developed by Robert Kaplan and David Norton (Balance Scorecard Institute [BSI], 2005). The scorecard approach is generally used for an entire organization but may also be used for a department or specific program. The balanced scorecard presents an organizational view from four perspectives: financial, customer, business processes, and learning and growth. The organization determines the objectives and metrics it should measure for each perspective necessary to support the larger vision or strategy. The financial perspective focuses on those financial areas relevant to the business or program such as profits, cost reduction, and cost-effectiveness data. The customer perspective could include determining exactly who are all the customers and their levels of satisfaction. The business process focuses on how well the business or program and its associated components are running. The learning and growth perspective may include identifying the organizational culture and training necessary to support the overall strategy. Figure 7 presents a generic view of a balanced scorecard.

Figure 7. Balanced Scorecard
(From BSI, 2005)
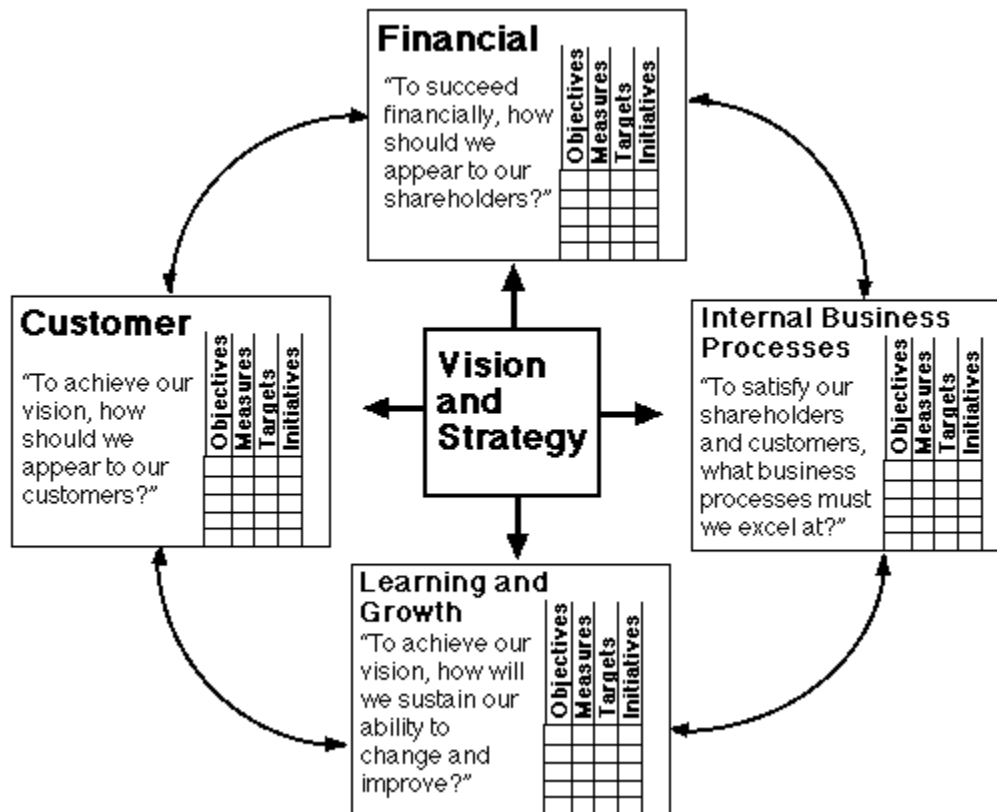
C. PROPOSED PILOT PROGRAM EVALUATION PLAN

As the results of the 360-degree feedback pilot program will be used to make decisions about further Navy-wide implementation, evaluation design must provide data for both judgment and improvement uses. Judgment uses will include impact assessments and cost-effectiveness analyses, while improvement uses will be guided by an implementation analysis. The design may also provide data that support knowledge uses for other training or policy decisions. The segmentation of the full pilot program into two distinct phases allows for assessment of Phase 2 impacts and implementation, which can then be used to make modifications to Phase 3. To provide more detailed evaluation information for ultimate decisions on program continuation, the overall program evaluation should include an impact assessment, an implementation evaluation, and a comprehensive cost-effectiveness analysis as a minimum. A balanced scorecard process may provide additional assistance by helping to identify all benefits and costs associated with the program.

51

1. Impact Assessment

The impact assessment should attempt to measure the actual effects of the program. The best method to assess impact is the experimental or quasi-experimental design. A control group should be designated for comparison to the Phase 2 experimental group. If there is not sufficient time to designate a control group for Phase 2, the most appropriate evaluation design would then be the pretest-posttest. The pretest-posttest allows for a summative evaluation of participant improvement based on scores both before and after the feedback intervention. The weakness of the pretest-posttest design is that it can only determine the program's gross effects, the total effects or changes in participants between measurements. The pretest-posttest design can not separate the program's net effects, those effects attributable specifically to the intervention, from the gross effects. The program's gross effects should be measured and then compared to the program's costs to produce an estimated cost-effectiveness analysis. The Navy must make a determination of whether or not the gross effects are sufficient to justify the costs of the program. If the program's gross effects are determined to be insufficient to justify the costs, the program should either be discontinued or modified to reduce costs. Modifications could include less frequent application of the survey or shortened surveys to assess only those areas identified for improvement in an individual's action plan. If the gross effects are assessed as sufficient, Phase 3 should be designed to allow more rigorous evaluation methods to provide an accurate cost-effectiveness analysis.

Quasi-experimental evaluation designs should be used in Phase 3. A matched control group that does not receive the feedback intervention should be designated for comparison with the experimental group. The experimental group should consist of two separate groups. One group should receive the feedback report and IDP only. The second group should receive the feedback report and IDP as well as coaching from the Commanding Officer or a designated mentor. The use of two experimental groups will allow for assessment of the impact of 360-degree feedback both with and without coaching. This quasi-experimental design will permit a more robust cost-effectiveness analysis of all aspects of the program. The assessment of costs and benefits is further developed in section C.3. of this chapter.

2. Implementation Analysis

An implementation analysis of all areas, effort, process, component, and treatment (see Figure 6), should be conducted for Phase 2 of the pilot program. A post-participation survey should be administered to all participants, including raters and ratees, to obtain their estimation of effort expended in the program and assessments of how well the program and its components are working. Analysis of NKO data on training course enrollment and completion can also inform the process and component evaluation. Treatment specification, which is also conducted in the impact assessment, should further attempt to determine which competencies have the greatest affect on leadership and which competencies are being identified most frequently for improvement in the IDPs and action plans.

Effort areas that should be measured are the NKO training course participation and completion rates and the use of a mentor or coach. Each of these areas is a significant component of the program and effort in these areas can directly affect program outcomes. Course participation and completion rates can be measured by monitoring NKO course registration and completion data and comparing these data to the courses recommended by the participant's IDP and action plan. Data on the use of a coach or mentor, including the number and frequency of mentoring sessions, is necessary for any attempt to determine a correlation between coaching and program impact.

Results of the effort evaluation can be used to inform the program process and component evaluation. The process and component evaluation should assess whether or not the parts of the program are working as designed or as desired. On-line training course participation and completion may be affected by internet connectivity. Course completion and use of a coach may both be affected by the time constraints of the participant's normal work load. The mentoring process may also be affected by the ratio of senior officers to junior officers in the command as well as possible personality conflicts that may prevent a member from seeking a mentor. Knowledge gained in the process and component areas should be used to determine if formal guidelines for NKO use and the mentoring process are warranted.

Treatment specification could be the most important segment of the implementation analysis as the results can be used to increase organizational knowledge and inform current policies in officer training and development. To enhance development efforts, the Navy should determine which of the leadership competency areas contribute most significantly to successful leadership within the Navy. The competencies should then be ranked in order of importance for leadership development. A ranked order of competencies could assist participants, Commanding Officers, and mentors in development and assessment of individual action plans. Action plans could be reviewed to ensure that participants are focusing efforts in those competencies determined to be most significant in leadership development. Focusing development on the most significant competencies could increase the amount of individual improvement between survey assessments and could increase the benefits and effectiveness of the overall program.

Additional treatment analysis should attempt to identify competencies that are consistently rated as deficient or proficient within specific organizational levels (e.g., Division Officer, Department Head, Executive Officer, Commanding Officer). Any consistencies noted could indicate a naturally occurring proficiency or deficiency within a specific organizational level. Knowledge of an organizational level's natural proficiencies and deficiencies could indicate an organizational need to incorporate specific training in those deficient competencies into the current CNL leadership training courses. Ultimately this analysis could lead to further customization of the survey instrument to target the specific development needs of each organizational level.

A final part of the treatment specification should be the validation of the survey instrument. As this instrument has not been used before, its reliability and validity can not be conclusively determined until used at length in the pilot program. While most sub-competencies in the pilot program are assessed by two to three questions each, others, such as those in the leading change core competency, are assessed by one question at most. A thorough assessment of the psychometric adequacy of the survey instrument should be conducted prior to its use in Phase 3 or in any future expansion of the program.

Overall results of the implementation analysis of Phase 2 should provide sufficient information to support design considerations for Phase 3. Results may suggest that only portions of the program need improvement or that major modifications might be necessary prior to any implementation of Phase 3.

3. Cost-Effectiveness Analysis

A comprehensive cost-effectiveness analysis should be undertaken when data collected are sufficient to permit evaluation. A quasi-experimental design, whether completed in Phase 2 or Phase 3, is necessary to determine program net effects, those effects that are directly attributable to the program. Program net effects should be compared to the program's total costs to assess the overall cost-effectiveness of the program.

The most significant costs of the program are the participant time requirements. The amount of time estimated for a rater to complete the pilot program survey is approximately fifteen minutes. The fifteen minutes required for a rater to complete a survey may appear inconsequential, but when measured across the entire organization, the time commitment can be quite substantial. For each feedback recipient, as many as ten surveys may be completed for each assessment period, one from self, and three each from supervisors, peers, and subordinates. Based on a survey completion time of fifteen minutes, and ten surveys per feedback recipient, 150 minutes may be expended to provide feedback to one individual. If the process occurs twice per year, 300 minutes are required to provide feedback to each individual. Approximately 125 man-years would be required to provide all officers, O-1 to O-5, with two feedback assessments per year. In addition to survey completion time, time to complete on-line courses, and time spent mentoring or coaching should also be included in the total time costs of the program. The annual programmed budget cost of a military officer should be used to quantify the personnel time cost. Other costs include the contractor cost of operating the 360-degree website and software program.

Determining program benefits includes, but is not limited too, measurement of actual program effects. Direct improvement attributable to the program is a benefit that can be weighed against program costs. However, the psychometric measure of benefits (i.e., the change is scores between assessments) may not capture all the psychosocial

55

benefits of using a 360-degree program for personal development. Other benefits may include improved organizational effectiveness, increased job satisfaction, improved retention and promotion rates, and increased knowledge that leads to improvements in organizational training and development. A balanced scorecard approach may be most useful in assessing all program benefits.

The balanced scorecard would assess the entire program in the four perspectives of financial, customer, business processes, and learning and growth. Figure 8 presents an abbreviated balanced scorecard for the pilot program, with possible benefits or objectives identified for each perspective; it offers an example of how the balanced scorecard could improve identification of program benefits.

Figure 8.   Elementary Balanced Scorecard for the Pilot Program

Perspective	Benefit or objective
Financial	- Return on investment (program impact vs. cost) - Increased retention beyond minimum service requirement - Increased promotion rates - Improved return on investment of other programs (NKO) - Improvement of other training resources (CNL leadership courses)
Customer	- Improved job satisfaction (both raters and ratees) - Greater awareness of self (ratees) - Personal development (improved feedback scores)
Business Process	- Increased use of NKO training resources - Increased use of coach or mentor - Improved organizational effectiveness
Learning and Growth	- Identification of organizational level proficiencies and deficiencies - Improved organizational training efforts to target proficiencies/deficiencies - Tailored surveys to target development needs of each organizational level - Impact of mentoring process

This basic balanced scorecard is not meant to provide an exhaustive list of the possible benefits of a 360-degree program, but is intended to illustrate how a balanced

scorecard approach may be a superior method of determining all benefits attributable to the program. In the absence of alternative programs for comparison, the Navy must be able to determine all benefits that accrue from using a 360-degree program to accurately assess those benefits against program costs. The balanced scorecard may provide information to support a more robust cost-effectiveness analysis to decide if the 360-degree program merits continuation or wider implementation.

D.   CONCLUSION

This chapter presents general guidelines for conducting a program evaluation. Evaluation designs are driven by the intended uses of the evaluation findings. If findings are to be used to make judgments about a program, then impact assessments and efficiency analyses are warranted. An implementation analysis should be conducted if findings are to be used to make improvements to a program or to increase organizational knowledge.

The evaluation results of the 360-degree feedback pilot program will be used to make both judgments about and improvements to the program and possibly to increase organizational knowledge. Based on these intended uses, a proposed program evaluation plan is presented. The plan includes an implementation analysis to identify areas for program improvement. An impact assessment and cost-effectiveness analysis, supported by a basic balanced scorecard, are included to guide data gathering for decisions regarding program continuation and wider implementation.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. CONCLUSIONS AND RECOMMENDATIONS

A. THESIS OVERVIEW

The objectives of this thesis were: 1) to identify research evidence on the effectiveness of 360-degree programs; 2) to identify best practices in using a 360-degree program; 3) to compare the Surface Warfare community's 360-degree pilot program to the research evidence; and 4) to provide a guideline for overall program evaluation. Chapter I presented the purpose of this thesis and discussed thesis scope, methodology, and expected benefits. Chapter II presented a brief history of 360-degree feedback use and research evidence on the effectiveness of 360-degree feedback as a development program. Chapter III discussed best practices of civilian and military programs that were used to complement the 360-degree feedback. Chapter IV described the Surface Warfare community's 360-degree pilot program and compared this program to the research evidence. Chapter V presented general program evaluation techniques and developed an evaluation guideline for use in evaluating the 360-degree feedback pilot program. This chapter provides overall conclusions and recommendations.

B. CONCLUSIONS

1. 360-degree Program Effectiveness

The use of 360-degree feedback as a development tool is based on the theory that ratings from multiple sources, such as supervisors, peers, and subordinates, are not similar and thus present the recipient with meaningful feedback data from the various sources' perspectives. Most research over the past decade has largely supported the theory of meaningful differences in multi-source ratings and found 360-degree programs to be effective development tools (Atwater et al., 1995; Walker and Smither, 1999; Hazucha et al., 1993; Reilly et al., 1996; Kluger and DeNisi, 1996). Recent research has introduced contradictory findings on the significance of dissimilarity between the ratings of various groups and questions past research findings on the magnitude of effectiveness of 360-degree programs (LeBreton et al., 2003; Scullen et al., 2000; Gregarus et al., 2003; Kluger and DeNisi, 1996). While the balance of the evidence largely supports a conclusion that 360-degree programs are effective development tools, most of that

evidence is based on studies conducted with non-experimental designs that were unable to separate the actual program effects from the effects of non-program factors that could have caused the improvement. Additional research on the effectiveness of 360-degree programs is warranted and organizations should fully evaluate potential costs and benefits prior to any large implementation of a 360-degree program.

2. 360-degree Program Best Practices

Several best practices to enhance the effectiveness of 360-degree programs were identified in the literature. One of the most beneficial practices identified is the use of an executive coach or feedback workshop to present feedback results, to assist with analysis of results and creation of development plans, and to conduct follow-up coaching sessions to ensure compliance with development plans. Three separate studies of 360-degree feedback coupled with executive coaching and feedback workshops found significant improvements in recipient feedback scores following the feedback intervention and coaching sessions (Thach, 2000; Luthans and Peterson, 2003; Seifert et al., 2002). Additionally, organizations that reported receiving high benefits from a 360-degree program were more likely to use internal rather than external coaches (Rogers et al., 2002). Other best practices identified were significant levels of training provided to all participants, the use of customized instruments targeted to specific organizational levels, and the use of 360-degree feedback for development vice performance appraisal purposes. The research supports the conclusion that organizations can significantly improve the effectiveness of their 360-degree programs by using an internal coach, by customizing surveys to specific organizational levels, and by using the program for development rather than appraisal purposes.

3. Surface Warfare Community 360-degree Pilot Program

The design of the 360-degree pilot program appears to be largely in line with the research evidence and the identified best practices. The program uses a single, customized survey for all participants to assess proficiency in the five core competencies of the Navy Leadership Competency Model. Feedback results are presented to the individual through the 360-degree program website. An Individual Development Plan (IDP) is also generated by the 360-degree software program that highlights deficient areas and provides links to electronic training resources, through the Navy Knowledge

Online (NKO) web portal, to help address those deficiencies. An executive coach is not assigned to each participant but the Commanding Officer and an undesignated mentor review IDP results and assist the recipient with the development of an action plan; thus the Command Officer and mentor act as internal coaches for the program. These findings support the conclusion that the 360-degree feedback pilot program should be an effective mechanism for personal development. Minor adjustments to the program are recommended and these are described in the recommendations section of this chapter.

4.   Program Evaluation

The design of a program evaluation is dependent on the intended uses of the findings. Findings of an evaluation generally serve three purposes: making judgments, identifying improvements, and increasing knowledge (Patton, 1997). Judgment oriented evaluations are most often used to make assessments about program effects and program continuation and are informed by impact assessments and cost-effectiveness analyses. Improvement oriented evaluations may be used to identify areas of a program that need adjustment and are usually informed by an implementation analysis. Knowledge oriented evaluations are largely conceptual and influence thinking and decisions about a specific program or policy. Knowledge evaluations are most often informed by implementation analyses but may also be informed by impact assessments and cost-effectiveness analyses.

Evaluation designs may be experimental, quasi-experimental, or non-experimental. Experimental designs randomly assign participants to an experimental group, the group that receives the treatment or intervention, and to a control group, the group that does not receive the treatment or intervention. Quasi-experimental designs are similar to experimental designs except that participants are not randomly selected and control groups are constructed by matching the control participants as closely as possible to the experimental participants. Non-experimental designs do not include control groups and are most often conducted by pretest-posttest measures on the experimental group. Experimental and quasi-experimental designs are superior to non-experimental designs as their inclusion of control groups allows for identification of the effects attributable solely to the treatment or intervention. A conclusion of this research is that a superior program evaluation would have an experimental or quasi-experimental design and would include

an impact assessment, an implementation analysis, and a cost-effectiveness analysis. Specific details for the conduct of these are outlined in Chapter V part C: Proposed Pilot Program Evaluation Plan.

C.   RECOMMENDATIONS

1.   Pilot Program Design

Based on a comparison of the Surface Warfare community's 360-degree pilot program with the research evidence and identified best practices, it is recommended that the pilot program use multiple instruments targeted to specific organizational levels (e.g., Division Officer, Department Head, Executive Officer, Commanding Officer), that the self-rating scores not be included in the average rating score for each competency, that the Navy consider using target scores rather than normative scores for identification of deficiencies, and that the mentoring process be more clearly defined and formalized.

Organizations that reported receiving high benefits from 360-degree feedback programs were more likely than those reporting low benefits to use survey instruments customized for each organizational level (Rogers et al., 2002). Additional research evidence suggests that feedback recipients do not attend equally to all sources of feedback (Gregarus et al., 2003), thus a single instrument may be presenting more feedback than would actually be used by the recipient. Many experts agree, though there is no empirical evidence offered to support the assertion, that a single instrument will suffer saturation after multiple uses over time and will lose its effectiveness as a development instrument. The pilot program survey should be customized to the level of the person being rated and to the competencies that raters typically observed (see Figure 2). Multiple survey instruments, customized to specific organizational levels and to the feedback that the recipient will actually attend to, present a superior method of preventing instrument saturation and of providing a continuum of developmental feedback throughout an individual's career progression.

Including the self-rating in the average of all ratings for each competency may potentially distort this overall score and affect the comparison with the normative score. If an individual's mean score for a particular competency is below the normative score,

that competency is designated as a development opportunity. Likewise, if the mean score is higher than the normative score, no improvements are suggested for that competency. Research has shown that self-ratings differ, sometimes significantly, from others' ratings (Atwater et al., 1995; Hazucha et al., 1993; Luthans and Peterson, 2003). Including the self-rating in the average may introduce an upward or downward bias and may cause inaccurate assessments of deficiency or proficiency in a competency.

While not specifically addressed in any of the 360-degree program effectiveness studies, the use of an "ideal" or target score for comparison with recipient feedback scores may be superior to using normative scores to identify development opportunities. The use of target scores may be especially beneficial in competencies that are determined to be more significant for successful leadership in the Navy. For example, if the Navy determined that "developing people" was an extremely significant competency for successful leadership, those who exceed the average, or normative, score would not have this competency identified as a development opportunity. However, an average score is not necessarily a non-achievable score for many people. Setting a target score higher than the normative score would cause more recipients to have this competency identified as a development opportunity and would help the Navy guide individual efforts toward further development of any identified critical competencies.

Research has shown that the use of a coach can significantly improve the effectiveness of a 360-degree program (Thach, 2002; Luthans and Peterson, 2003; Seifert et al., 2002). The Surface Warfare pilot program dictates that the Commanding Officer review 360-degree program IDPs and action plans with each individual during the mid-term counseling session. During the follow-up assessment six months later, a mentor is used instead of the Commanding Officer. It is unclear if the mentor is selected by the command or by the individual. It is also not known if the mentor participates in any way in the mid-term 360-degree assessment. The mentoring process should be clarified in the 360-degree program instructions to include selection and participation in all assessments and guidance for development of broader reaching individual action plans. A formal mentoring process will ensure that each participant clearly understands this process and that each has access to an internal coach throughout the process to assist with "whole person" development.

63

2. Pilot Program Evaluation

Based on the research evidence, it is recommended that a quasi-experimental design be used to evaluate the Surface Warfare 360-degree pilot program. Program evaluation should include an impact assessment, an implementation analysis, and a cost-effectiveness analysis as outlined in Chapter V of this thesis.

An impact assessment requires construction of a matched control group for Phase 2 to determine the effects that can be attributed solely to the 360-degree program. If time does not permit designation of a control group for Phase 2, the primary alternative is the non-experimental pretest-posttest measurement to determine whether or not the program produces positive effects, however this design can not determine causality because it is non-experimental and can not separate the effects of the program from the effects of other factors external to the program.

It is strongly recommended that a control group be designated for Phase 2 to allow a greater breadth of impact assessments in Phase 3. Research evidence suggests that most improvement occurs between the first and second application of a feedback instrument and that this improvement can be sustained with less frequent follow-up applications (Reilly et al., 1996; Walker and Smither, 1999). If a control group is used in Phase 2, actual program effects between the first and second assessment can be determined. The Phase 2 experimental group could continue the program as the Phase 3 experimental group and could then be used to assess the sustainability of improvements and to look for indicators of instrument saturation. The Phase 3 experimental group could be divided into two groups. The first experimental group would continue the process as currently designed with reapplication of the instrument every six months. Individuals in this group could receive as many as six applications of the instrument over the course of both Phase 2 and Phase 3. Any reduction in improvement levels could signal instrument saturation. The second experimental group in Phase 3 would receive only one 360-degree assessment, approximately one year after their last Phase 2 assessment. This group's results could indicate whether or not the improvements are sustainable with less frequent reapplication of the instrument. The results of both groups could be used to make decisions about how frequently the instrument should be applied

to maintain improvements and how many times the instrument can be used before its developmental impact degrades.

An additional test could be performed with the experimental groups to determine the impact of the coaching/mentoring process. Participants could be divided into a group that receives feedback only and a group that receives feedback and coaching/mentoring to further isolate the effects of the feedback from that of the coaching process.

If a control group is not designated until Phase 3, the impact assessments described above will not be possible. New experimental participants would be necessary for Phase 3 to determine actual program effects as Phase 2 participants will have previously received the intervention and will likely have made improvements as a result of the intervention. Based on the research, Phase 2 participants would not show as much improvement as would new experimental participants, therefore Phase 3 assessment results could potentially be contaminated by using Phase 2 participants in Phase 3.

The implementation analysis should be informed by a post-participation survey. The survey should be administered to all participants, including raters and ratees, to obtain their estimation of effort expended in the program and assessments of how well the program and its components, such as mentoring and NKO training, are working. Analysis of NKO data on training course enrollment and completion can also inform the implementation analysis. The survey should seek to determine participant satisfaction with the program and to identify areas suggested for improvement.

Another focus of the implementation analysis should be the Navy Leadership Competency Model. Five core competencies with twenty-five associated sub-competencies are listed; however there is no indication as to which competencies contribute most significantly to successful leadership in the Navy. For development purposes, the Navy should rank order the competencies according to their impact on successful leadership. A ranked order of competencies would assist individuals and Commanding Officers/mentors in developing action plans that target improvements in those competencies deemed most significant.

Survey results for each organizational level should also be analyzed to determine if there are competencies that are consistently rated as deficient for a particular group.

Any consistent deficiencies noted could indicate a need to incorporate specific training in that competency into current Navy Leadership Development courses. For example, if Ensigns were consistently rated as deficient in financial management, the Navy could incorporate specific financial management training into the Basic Officer Leadership course to target this deficiency.

When pilot program data become available, a comprehensive cost-effectiveness analysis should be conducted. A determination must be made regarding whether program benefits outweigh the costs to achieve those benefits. While costs, such as participant time and contractor administration, can be readily quantified, benefits include more than just improved 360-degree scores and can be quite difficult to quantify. A balanced scorecard approach, as outlined in Chapter V of this thesis, is recommended as a more comprehensive process of identifying and quantifying all costs and benefits associated with the 360-degree program. An accurate assessment of all costs and benefits is necessary to inform decisions about program continuation and wider implementation.

D.   RECOMMENDATIONS FOR FUTURE RESEARCH

This thesis presents a conceptual framework for evaluating the Surface Warfare community's 360-degree pilot program. Using the guideline presented in this thesis, future research should be conducted in the following areas:

- Validation of the psychometric adequacy of the survey instrument.

- Statistical analysis of pilot program survey results to determine program effects.

- Analysis of the Navy Leadership Competency Model to determine which competencies contribute most significantly to successful leadership in the Navy.

- Analysis of pilot program survey results to determine if specific organizational levels are consistently rated as deficient in any competencies. If deficiencies exist, conduct an analysis of how best to

incorporate specific training for these deficiencies into current Navy Leadership Development training.

- Comprehensive cost-effectiveness analysis of the pilot program.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX   PILOT PROGRAM SURVEY QUESTIONS

Accomplishing Mission

1. Seeks ideas for improvements.
2. Knowledgable of current events.
3. Aware of external issues impacting command mission.
4. Committed to the Navy.
5. Clearly defines goals for the command.
6. Clearly plans for the future of the command.
7. Supports the chain of command.
8. Communicates the command vision.
9. Works to achieve the command vision.
10. Provides clear direction on command mission.
11. Works to achieve the command mission.
12. Holds self accountable for actions.
13. Holds others accountable for actions.
14. Able to make a decision.
15. Considers risk during daily execution.
16. Solves problems.
17. Clearly defines subordinate's job.
18. Clearly defines subordinate's responsibility.

Resource Stewardship

19. Budgets for command needs.
20. Uses funds as budgeted.
21. Uses technology to improve productivity.
22. Effectively deals with personnel.
23. Completes projects on time.
24. Completes projects within budget.
25. Uses continuous improvement methods.
26. Uses planning to manage resources
27. Acts according to plan.
28. Uses resources well.
29. Develops subordinates professionally.
30. Promotes health and fitness.

Working with People

31. Mentors subordinates.

32. When speaking, gets the point across.

33. Speaks clearly.

34. Adjusts well to changes

35. Listens to other's ideas.

36. Encourages safe behavior.

37. Supports the team.

38. Supports the navy culture.

39. Communicates well in writing.

40. Relates well with others.

41. Is a good listener.

42. Is a team player.

43. Others like working with him/her.

Leading People

44. Does not abuse authority.

45. Helps subordinates with personal problems

46. Helps subordinates prepare for advancement.

47. Resolves issues among subordinates.

48. Respects cultural differences.

49. Respects gender differences.

50. Acts professionally.

51. Gets subordinates to work as a team.

52. Leads well in a crisis.

53. Prepares subordinates for combat.

54. Delegates effectively.

55. Is honest.

56. Leads by example.

57. Acts according to his/her words.

58. Inspires confidence.

59. Motivates me.

60. Provides positive feedback.

61. Provides positive reinforcement.

Leading Change

62. Develops unique and effective solutions.

63. Acts appropriately.

64. Strives to improve as a person.

65. Strives to improve professionally.

66. Is skillful in his/her job.

67. Uses technology at work.

68. Can be trusted.

LIST OF REFERENCES

Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology, 47*(2), 349.

Atwater, L., Roush, P., Fischthal, A. (1995). The influence of upward feedback on self- and follower ratings of leadership. *Personnel Psychology, 48*(1), 35.

Brett, J. F., & Atwater, L. E. (2001). 360 degree feedback: Accuracy, reactions, and perceptions of usefulness. *The Journal of Applied Psychology, 86*(5), 930.

Brutus, S., & Derayeh, M. (2002). Multisource assessment programs in organizations: An insider's perspective. *Human Resource Development Quarterly, 13*(2), 187.

Brutus, S., Fleenor, J. W., & London, M. (1998). Elements of effective 360-degree feedback. In W. W. Tornow, M. London, & CCL associates, *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (1st ed.). San Francisco: Jossey-Bass.

BUPERS Instruction 1610.10 (1995). Navy performance evaluation and counseling system. Retrieved April 1, 2005 from http://buperscd.technology.navy.mil/bup_updt/508/Instructions/161010/161010.html

Cascio, W. F. (1998). *Applied psychology in human resource management* (5th ed.). Upper Saddle River: Prentice Hall.

CCL's 2005 Assessment and Development Resources catalog. Retrieved April 15, 2005 from http://www.ccl.org/CCLCommerce/pdf/assessments/ADRSection2004.pdf

Chappelow, C. (2003). Dividends & interest -- news flash: 360-degree feedback is alive and well. *Leadership in Action, 23*(2), 22.

Church, A. H., & Bracken, D. W. (1997). Advancing the state of the art of 360-degree feedback. *Group & Organization Management, 22*(2), 149.

Conway, J., & Huffcutt, A. (1997). Psychometric properties of multi-source performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*, 331.

Dalton, M. A. (1998). Best practices: Five rationales for using 360-degree feedback in organizations. In W. W. Tornow, M. London, & CCL associates, *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (1st ed.). San Francisco: Jossey-Bass.

Edwards, M. R., & Ewen, A. J. (1986). *360° feedback : The powerful new model for employee assessment & performance improvement*. New York: AMACOM.

Eichinger, R. W., & Lombardo, M. M. (2003). Knowledge summary series: 360-degree assessment. *HR.Human Resource Planning, 26*(4), 34.

Ghorpade, J. (2000). Managing five paradoxes of 360-degree feedback. *The Academy of Management Executive, 14*(1), 140.

Greguras, G. J., Ford, J. M., & Brutus, S. (2003). Manager attention to multisource feedback. *The Journal of Management Development, 22*(4), 345.

Greguras, G. J., & Robie, C. (1998). A new look at within-source interrater reliability of 360-degree feedback ratings. *The Journal of Applied Psychology, 83*(6), 960.

Harris, M. M., & Schaubroeck, J. (1988) A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*(1), 43.

Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management, 32*(2,3), 325.

Hedge, J. W., Borman, W. C., & Birkeland, S. A. (2001). History and development of multisource feedback as a methodology. In D. Bracken, C. W. Timmereck, & A. Church (Eds.), *Handbook of multisource feedback*. San Francisco: Jossey-Bass.

James, R, L., Demaree, G, R., Wolf, & Gerrit. (1984). Estimating within-group interrater reliability with and without response bias. *The Journal of Applied Psychology, 69*(1), 85.

Kluger, N, A., DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254.

LeBreton, J. M., Burgess, J. R. D., Kaiser, R. B., Atchley, E. K., & James, L. R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods, 6*(1), 80.

Lepsinger, R., & Lucia, A. D. (1997). *The art and science of 360° feedback*. San Francisco: Jossey-Bass.

London, M., Smither, J. W., & Adsit, D. J. (1997). Accountability: The achilles' heel of multisource feedback. *Group & Organization Management, 22*(2), 162.

London, M., & Tornow, W. W. (1998). 360-degree feedback – more than a tool. In W. W. Tornow, M. London, & CCL associates, *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (1st ed.). San Francisco: Jossey-Bass.

Luthans, F., & Peterson, S. J. (2003). 360-degree feedback with systematic coaching: Empirical analysis suggests a winning combination. *Human Resource Management, 42*(3), 243.

Martineau, J. W. (1998). Using 360-degree surveys to assess change. In W. W. Tornow, M. London, & CCL associates, *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (1st ed.). San Francisco: Jossey-Bass.

Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1988). Trait, rater, and level effects in 360-degree performance ratings. *Personnel Psychology, 51*(3), 557.

Naval Administrative Message 069/04. Navy Leadership Development Program for Enlisted Sailors. Retrieved April 1, 2005 from http://www.persnet.navy.mil/navadmin/nav04/nav04069.txt

Naval Personnel Development Command. (2005, March 9). Retrieved April 15, 2005 from https://www.npdc.navy.mil/

Navy Knowledge Online. (2004, October 29). Retrieved April 15, 2005 from https://wwwa.nko.navy.mil/portal/splash/index.jsp

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd. ed.) Thousand Oaks: Sage.

Pfau, B., Kay, I., Nowack, K. M., & Ghorpade, J. (2002). Does 360-degree feedback negatively affect company performance? *HRMagazine, 47*(6), 54.

Posavac, E. J., & Carey, R. G. (1989). *Program evaluation: Methods and case studies* (3rd ed.). Englewood Cliffs: Prentice Hall.

Reilly, R, R., Smither, J. W., & Vasilopoulos, N. L. (1996). A longitudinal study of upward feedback. *Personnel Psychology, 49*(3), 599.

Rogers, E., Rogers, C. W., & Metlay, W. (2002). Improving the payoff from 360-degree feedback. *HR.Human Resource Planning, 25*(3), 44.

Rossi, P. H., & Freeman, H. E. (1989). *Evaluation: A systematic approach* (4th ed.). Newbury Park: Sage.

Scullen, S. E., Goff, M., & Mount, M. K. (2000). Understanding the latent structure of job performance ratings. *The Journal of Applied Psychology, 85*(6), 956.

Seifert, C. F., Yukl, G., & McDonald, R. A. (2003). Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *The Journal of Applied Psychology, 88*(3), 561.

Spinner, J. (2005). Submarine squadron 20 to conduct 360-degree review of commanding officers. *Navy newsstand*. Story number: NNS050329-18. Retrieved March 30, 2005 from http://www.news.navy.mil/search/display.asp?story_id=17704.

Thach, E. C. (2002). The impact of executive coaching and 360 feedback on leadership effectiveness. *Leadership & Organization Development Journal, 23*(3/4), 205.

Tornow, W. W. (1998). Forces that affect the 360-degree feedback process. In W. W. Tornow, M. London, & CCL associates, *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (1st ed.). San Francisco: Jossey-Bass.

Van Velsor, E. (1998). Designing 360-degree feedback to enhance involvement, self-determination, and commitment. In W. W. Tornow, M. London, & CCL associates, *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (1st ed.). San Francisco: Jossey-Bass.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2002). Moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: Confounding construct-level convergence and rating difficulty. *Journal of Applied Psychology, 87*(2), 345.

Walker, A. G., & Smither, J. W. (1999). A five-year study of upward feedback: What managers do with their results matters. *Personnel Psychology, 52*(2), 393.

What is the balanced scorecard? The Balanced Scorecard Institute (2005, April) Retrieved April 21, 2005 from http://www.balancedscorecard.org/basics/bsc1.html

INITIAL DISTRIBUTION LIST

1.   Defense Technical Information Center
        Ft. Belvoir, Virginia

2.   Dudley Knox Library
        Naval Postgraduate School
        Monterey, California

3.   LCDR James Pfautz
        Center for Naval Leadership
        Norfolk, Virginia

4.   CDR David Bossert
        Naval Personnel Command
        Millington, Tennessee

5.   Mr. Jeff Munks
        Naval Postgraduate School
        Monterey, California

6.   Professor Susan Hocevar
        Naval Postgraduate School
        Monterey, California

7.   Professor Alice Crawford
        Naval Postgraduate School
        Monterey, California

8.   CDR Sue Bosque
        Naval Postgraduate School
        Monterey, California